



US009447404B2

(12) **United States Patent**  
**Breuer et al.**

(10) **Patent No.:** **US 9,447,404 B2**  
(45) **Date of Patent:** **Sep. 20, 2016**

(54) **METHOD FOR THE BIOCATALYTIC CYCLIZATION OF TERPENES AND CYCLASE MUTANTS EMPLOYABLE THEREIN**

USPC ..... 435/233, 252.3, 254.2, 320.1  
See application file for complete search history.

(56) **References Cited**

(71) Applicant: **BASF SE**, Ludwigshafen (DE)

FOREIGN PATENT DOCUMENTS

(72) Inventors: **Michael Breuer**, Darmstadt (DE);  
**Bernhard Hauer**, Fußgönheim (DE);  
**Dieter Jendrossek**, Tübingen (DE);  
**Gabrielle Siedenburg**, Siedenburg (DE);  
**Juergen Pleiss**, Asperg (DE);  
**Demet Sirim**, Stuttgart (DE); **Silvia Fadenrecht**, Stuttgart (DE)

WO WO-2010/139719 A2 12/2010

OTHER PUBLICATIONS

(73) Assignee: **BASF SE**, Ludwigshafen (DE)

(\* ) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

Devos et al., *Proteins: Structure, Function and Genetics*, 2000, vol. 41: 98-107.\*  
Whisstock et al., *Quarterly Reviews of Biophysics* 2003, vol. 36 (3): 307-340.\*  
Witkowski et al., *Biochemistry* 38:11643-11650, 1999.\*  
Kisselev L., *Structure*, 2002, vol. 10: 8-9.\*  
“Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB). *Enzyme Supplement 5 (1999)*”, *Eur. J. Biochem.*, 1999, vol. 264, pp. 610-650.  
Daum et al., “Genes and Enzymes Involved in Bacterial Isoprenoid Biosynthesis”, *Current Opinion in Chemical Biology*, 2009, vol. 13, pp. 180-188.  
Devos et al., “Practical Limits of Function Prediction”, *Proteins: Structure, Function and Genetics*, 2000, vol. 41, pp. 98-107.  
Kisselev, L., “Polypeptide Release Factors in Prokaryotes and Eukaryotes: Same Function, Different Structure”, *Structure*, 2002, vol. 10, pp. 8-9.  
Neumann et al., “Purification, Partial Characterization and Substrate Specificity of a Squalene Cyclase from *Bacillus acidocaldarius*”, *Biol. Chem.*, 1986, vol. 367, pp. 723-729.  
Seo et al., “The Genome Sequence of the Ethanologenic Bacterium *Zymomonas mobilis* ZM4” *Nature Biotechnology*, 2005, vol. 23, No. 1, pp. 63-68.  
Whisstock et al., “Prediction of Protein Function From Protein Sequence and Structure”, *Quarterly Reviews of Biophysics*, 2003, vol. 36, No. 3, pp. 307-340.  
Witkowski et al., “Conversion of a  $\beta$ -Ketoacyl Synthase to a Malonyl Decarboxylase by Replacement of the Active-Site Cysteine with Glutamine”, *Biochemistry*, 1999, vol. 38, pp. 11643-11650.

(21) Appl. No.: **14/560,263**

(22) Filed: **Dec. 4, 2014**

(65) **Prior Publication Data**

US 2015/0104851 A1 Apr. 16, 2015

**Related U.S. Application Data**

(62) Division of application No. 13/297,798, filed on Nov. 16, 2011, now Pat. No. 8,932,839.

(60) Provisional application No. 61/414,434, filed on Nov. 17, 2010, provisional application No. 61/499,228, filed on Jun. 21, 2011, provisional application No. 61/540,028, filed on Sep. 28, 2011.

(51) **Int. Cl.**

**C12N 9/10** (2006.01)  
**C12N 1/19** (2006.01)  
**C12P 5/02** (2006.01)  
**C12N 9/90** (2006.01)  
**C12N 9/88** (2006.01)  
**C12P 7/02** (2006.01)  
**C12P 5/00** (2006.01)

(52) **U.S. Cl.**

CPC . **C12N 9/90** (2013.01); **C12N 9/88** (2013.01);  
**C12P 5/007** (2013.01); **C12P 7/02** (2013.01);  
**C12Y 504/99017** (2013.01); **Y02P 20/52**  
(2015.11)

(58) **Field of Classification Search**

CPC ..... C12N 9/88; C12N 9/90; C12P 5/007;  
C12Y 504/9901

\* cited by examiner

*Primary Examiner* — MD. Younus Meah

(74) *Attorney, Agent, or Firm* — Drinker Biddle & Reath LLP

(57) **ABSTRACT**

The present invention relates to novel mutants with cyclase activity and use thereof in a method for biocatalytic cyclization of terpenes, such as in particular for the production of isopulegol by cyclization of citronellal; a method for the preparation of menthol and methods for the biocatalytic conversion of further compounds with structural motifs similar to terpene.

**14 Claims, 4 Drawing Sheets**

1	MGIDRMNSLS	RLLMKKIFGA	EKTSYKPASD	TIIGTDTLKR	PNRRPEPTAK
51	VDKTI FKTMG	NSLNNTLVSA	CDWLI GQOKP	DGHWVGAVES	NASMEA EWCL
101	ALWFLGLEDH	PLRPRLGNAL	LEMQREDGSW	GVYFGAGNGD	INATVEAYAA
151	LRSLGYSADN	PVLKKA AAWI	AEKGLKNI R	VFTRYWLALI	GEWPWEKTPN
201	LPPEI I WFPD	NFVFSIYNFA	QWARATMVPI	AILSARRPSR	PLRPQDRLDE
251	LFPEGRARFD	YELPKKEGID	LWSQFRTTD	RGLHWVQSNL	LKRNSLREAA
301	IRHVLEWII R	HQDADGGWGG	IQPPWVYGLM	ALHGEGYQLY	HPVMAKALSA
351	LDDPGWRHDR	GESSWIQATN	SPVWDTMLAL	MALKDAKAED	RFTPEMDKAA
401	DWLLARQVKV	KGDWSIKLPD	VEPGGWAFEY	ANDRYPDTDD	TAVALIALSS
451	YRDKEEWQKK	GVEDAITRGV	NWLIAMQSEC	GGWGA FDKDN	NRSILSKIPF
501	CDFGESIDPP	SVDVTAHVLE	AFGTLGLSRD	MPVIQKAIDY	VRSEQEAEGA
551	WFGRWGVNYI	YGTGAVLPAL	AAIGEDMTQP	YITKACDWLV	AHQQEDGGWG
601	ESCSSYMEID	SIGKGPTTFS	QTAWALMGLI	AANRPEDYEA	IAKGCHYLID
651	RQEQDGSWKE	EEFTGTGFPG	YGVGQTIKLD	DPALSKRLLQ	GAELSR AFML
701	RYDFYRQFFP	IMALSRAERL	IDLNN		

Fig. 1a

1	atgggtattg	acagaatgaa	tagcttaagt	cgcttgtaa	tgaagaagat
51	tttcggggct	gaaaaaacct	cgtataaac	ggcttccgat	accataatcg
101	gaacgggatac	cctgaaaaga	ccgaaccggc	ggcctgaacc	gacggcaaaa
151	gtcgacaaa	cgatattcaa	gactatgggg	aatagctga	ataataccct
201	tgtttcagcc	tgtgactggt	tgatcggaca	acaaaagccc	gatggctatt
251	gggtcggctgc	cggtggaatcc	aatgcttcga	tggaaagcaga	atgggtgctg
301	gccttggtgt	ttttgggtct	ggaagatcat	ccgcttcgct	caagattggg
351	caatgctctt	ttggaaatgc	agcgggaaga	tggctcttgg	ggaggtctatt
401	tcggcgctgg	aaatggcgat	atcaatgcc	cggttgaagc	ctatgcggcc
451	ttgcggctct	tggggatattc	tgccgataat	cctgttttga	aaaaagcggc
501	agcatggatt	gctgaaaaag	gcggattaaa	aaatatccgt	gtctttacc
551	gttattggct	ggcggtgatc	ggggaatggc	cttgggaaaa	gaccctaac
601	cttccccctg	aaattatctg	gttccctgat	aattttgtct	tttcgattta
651	taattttgcc	caatggggcgc	gggcaaacat	ggtgcccatt	gctattctgt
701	ccgcgagacg	accaagccgc	ccgctgcgoc	ctcaagaccg	attggatgaa
751	ctgtttccag	aaggccgcgc	tcgctttgat	tatgaattgc	cgaaaaaaga
801	agccatcgat	ctttggctgc	aatttttccg	aaccactgac	cggttttac
851	attgggttca	gtccaatctg	ttaaagcgea	atagcttgcg	tgaagccgct
901	atccgctcatg	ttttggaatg	gattatccgg	catcaggatg	ccgatggcgg
951	ttgggttga	attcagccac	cttgggtcta	tggtttgatg	gcgtttacatg
1001	gtgaaggcta	tcagctttat	catccggtga	tggccaaggc	tttgctggct
1051	ttggatgatc	ccggttggcg	acatgacaga	ggcgagtctt	cttggataca
1101	ggccaccaat	agtcgggtat	gggatacaat	gttggccttg	atggcgttaa
1151	aagacgccaa	ggccgaggat	cgttttacgc	cggaaaatgga	taaggccgcc
1201	gattggcctt	tggctcgaca	ggtcaaagtc	aaagggcatt	ggtcaatcaa
1251	actgcccgat	gttgaaccgg	gtggatgggc	atttgaatat	gccaatgatc
1301	gctatcccga	taccgatgat	accgcccgtcg	ctttgatcgc	cctttcctct
1351	tatcgtgata	aggaggagtg	gcaaaagaaa	ggcgttgagg	agccattac
1401	ccgtggggtt	aattggttga	tcgccatgca	aagcgaatgt	ggcggttggg
1451	gagcccttga	taaggataat	aacagaagta	tcctttccaa	aatcctttt
1501	tgtgatttcg	gagaatctat	tgatccgcct	tcagtcgatg	taaccggcga
1551	tgttttagag	gcctttggca	ccttgggact	gtcccggat	atgccggtca
1601	tccaaaaagc	gatcgactat	gtccgctccg	aacaggaagc	cgaagggcgg
1651	tggtttggtc	gttggggcgt	taattatatac	tatggcaccg	gtgcggttct
1701	gcctgctttg	gcggcgatcg	gtgaagatat	gaccagcct	tacatcacca
1751	aggcttgcg	ttggctggtc	gcacatcagc	aggaagacgg	cggttggggc
1801	gaaagctgct	cttctatatac	ggagattgat	tccattggga	agggcccacc
1851	cagcgcgtcc	cagactgctt	gggctttgat	ggggttgatc	gcggccaatc
1901	gtcccgaaga	ttatgaagcc	attgccaagg	gatgccatta	tctgatgat
1951	cgccaagagc	aggatggtag	ctggaaagaa	gaagaattca	ccggcaccgg
2001	attccccggt	tatggcgtgg	gtcagacgat	caagttggat	gatccggctt
2051	tatcgaacg	attgcttcaa	ggcgtgaac	tgtcacgggc	gtttatgctg
2101	cgttatgat	tttatcggca	attctccc	attatggcgt	taagtccggc
2151	agagagactg	attgatttga	ataattga		

Fig. 1b

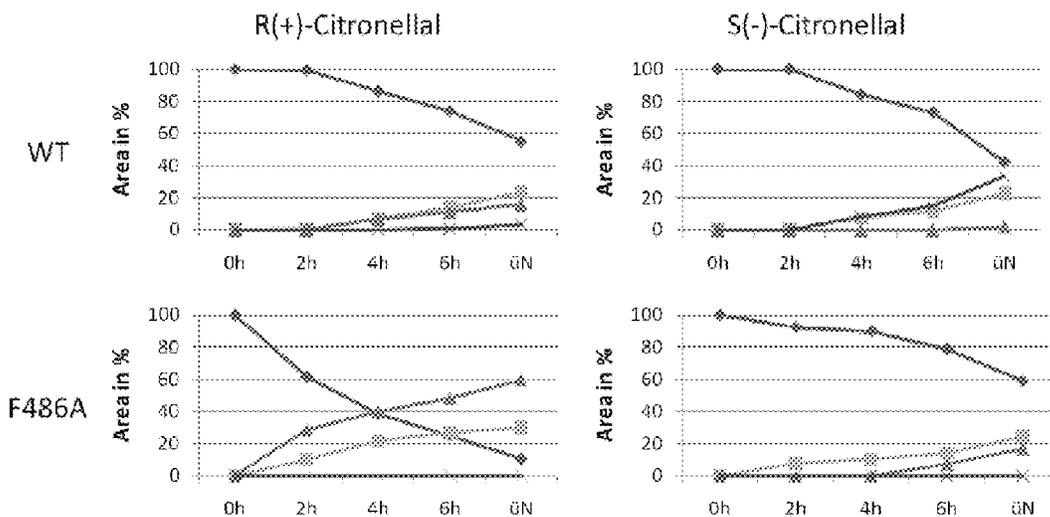


Fig. 2

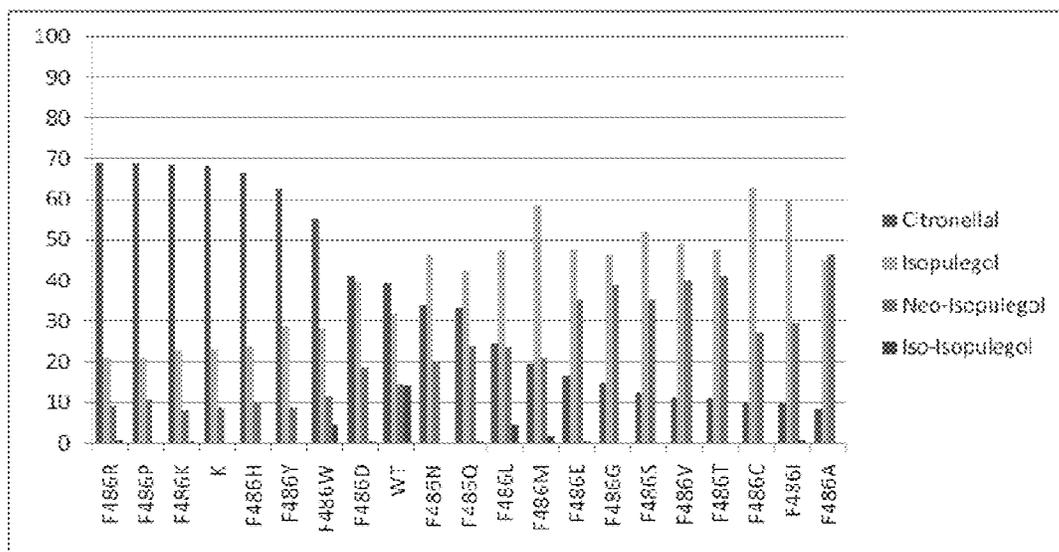


Fig.3

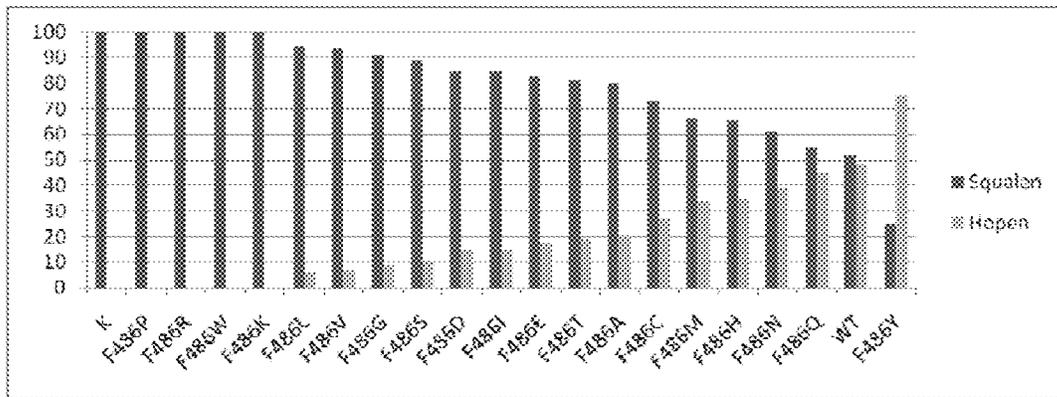


Fig. 4

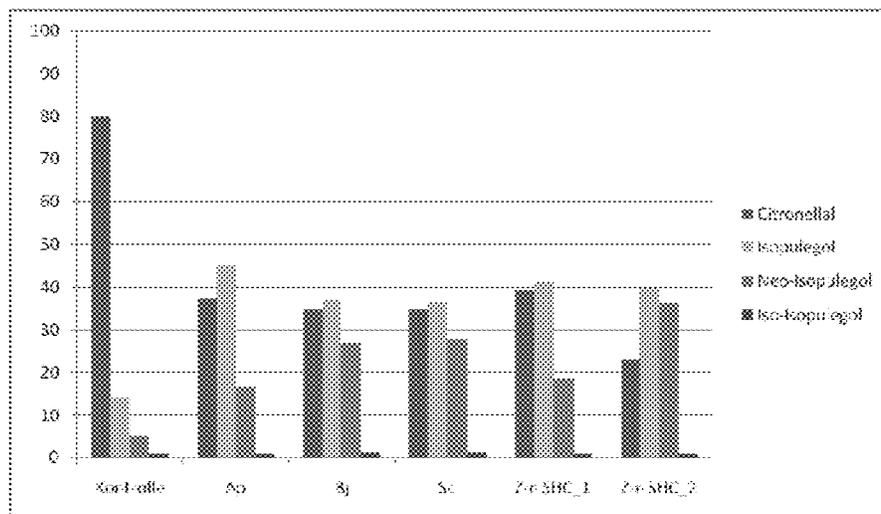


Fig.5

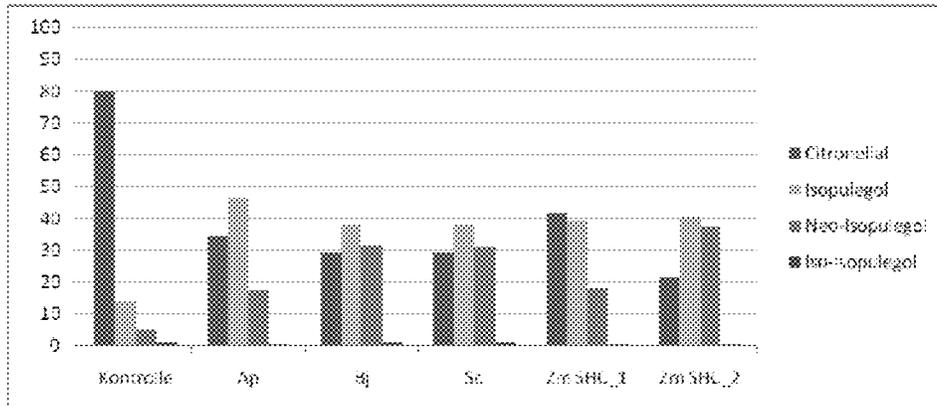


Fig.6

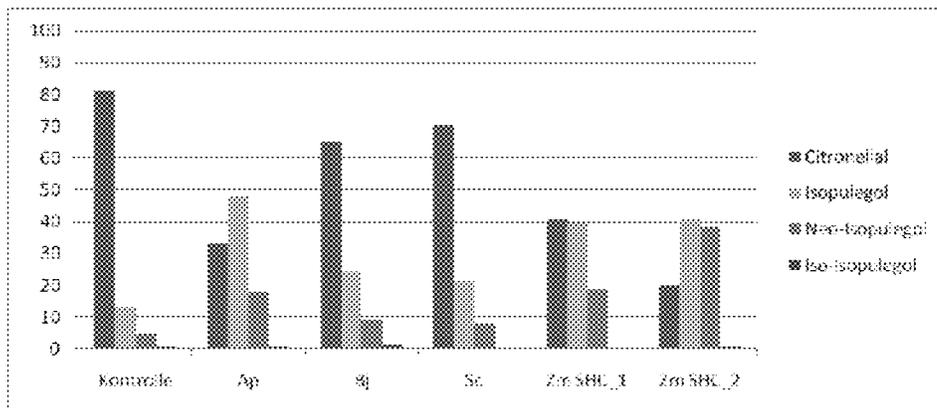


Fig.7

1

**METHOD FOR THE BIOCATALYTIC  
CYCLIZATION OF TERPENES AND  
CYCLASE MUTANTS EMPLOYABLE  
THEREIN**

RELATED APPLICATIONS

This application is a divisional of U.S. application Ser. No. 13/297,798, filed Nov. 16, 2011, now U.S. Pat. No. 8,932,839, which claims the benefit under 35 U.S.C. §119(e) of U.S. Provisional Application 61/414,434, filed Nov. 17, 2010; U.S. Provisional Application 61/499,228, filed Jun. 21, 2011; and U.S. Provisional Application 61/540,028, filed Sep. 28, 2011. The entire content of each aforementioned application is hereby incorporated by reference in its entirety.

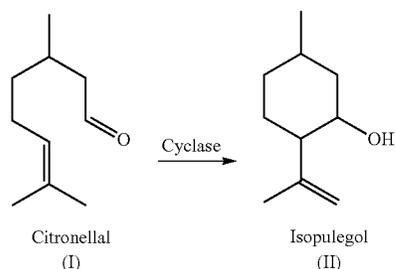
SUBMISSION OF SEQUENCE LISTING

The Sequence Listing associated with this application is filed in electronic format via EFS-Web and hereby incorporated by reference into the specification in its entirety. The name of the text file containing the Sequence Listing is Sequence\_List\_074012\_00194\_01. The size of the text file is 1,428 KB, and the text file was created on Dec. 2, 2014.

The present invention relates to novel methods for cyclizing terpenes using cyclases and to novel mutants with cyclase activity and use thereof in a method for biocatalytic cyclization of terpenes, such as in particular for the production of isopulegol by cyclization of citronellal; a method for the preparation of menthol and methods for the biocatalytic conversion of further compounds with structural motifs similar to terpene.

BACKGROUND OF THE INVENTION

Isopulegol of formula (II) (2-isopropenyl-5-methyl-cyclohexanol) is a terpene that is used as an aroma compound, to generate "flower notes". Moreover, it is an intermediate in the synthesis of menthol from citral.



Isopulegol isomers occur in nature in a large number of essential oils. As isopulegol is formed relatively easily from citronellal, the compound of formula (I) (3,7-dimethyloct-6-en-1-al), it often occurs accompanying citronellal or is formed during extraction of the essential oil. Isopulegol, which is produced industrially from (+)-citronellal, is as a rule a mixture of different isomers with a high proportion of (-)-isopulegol.

The industrial production of isopulegol is mainly carried out by the chemical cyclization of (+)-citronellal. Originally 80-85% pure raw material obtained from citronella oil was used. Since the 1990 s this has increasingly been replaced with the optically purer (+)-citronellal (97.5%) from the so-called Takasago process. Here, geranyldiethylamine is isomerized asymmetrically to (+)-citronellal using an Rh-BINAP-complex catalyst (Rh-complex with 2,2'-bis-(diphenylphosphino)-1,1'-binaphthyl).

2

The chemical synthesis of isopulegol starting from citronellal has been described many times. (+)-Citronellal can be cyclized using a copper-chromium catalyst, zinc bromide, alkylaluminum chloride, a rhodium complex, a solid acid-base catalyst, zeolite or silica gel. In recent times the silica gel method has increasingly been superseded by the method with zinc bromide, as the latter has higher selectivity.

The cyclization of terpenes with the aid of special cyclases is generally known. For example, in nature squalene is cyclized by a squalene-hopene cyclase (SHC) to the pentacyclic hopene.

The gene and protein sequences of squalene-hopene cyclase derived from the bacterium *Zymomonas mobilis* (Zm-SHC) are known (Genpept Accession No AAV90172 2004 and Nat Biotechnol 2005, 23:63-68, cf. SEQ ID NO: 1 and 2).

In international application PCT/EP2010/057696 (WO2010139719 A2), to the complete disclosure of which reference is expressly made herein, polypeptides are proposed as biocatalysts for the cyclization of homofarnesol to ambroxan.

The biosynthesis of numerous monoterpenes in the corresponding production organisms has already been elucidated. Frequently this involves cyclization of linear precursor molecules by highly specific biocatalysts. The precursors are generally esters of linear terpene alcohols and diphosphoric acid. One typical example of such a precursor is geranyl pyrophosphate. The pyrophosphate group is eliminated from the molecule enzymatically, and is subsequently hydrolyzed into two phosphate ions. On the other side, a carbocation is formed, which is then able to undergo further intramolecular reaction and which recombines to form a cyclic monoterpene, with elimination of a proton, for example (Curr. Opin. Chem. Biol. 2009, 13: 180-188).

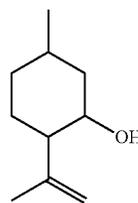
A problem to be solved by the present invention, furthermore, was to find an alternative to the known chemical cyclization methods for terpenes, allowing terpene compounds to be cyclized by means of enzymatic catalysis, such as the linear citronellal to be cyclized to isopulegol, for example.

The problem to be solved by the present invention was furthermore to provide novel biocatalysts that can be used for the cyclization of terpenes, for example of citronellal with formation of isopulegol.

SUMMARY OF THE INVENTION

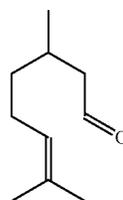
The above first problem is solved by a method of production of isopulegol of general formula (I)

(I)



comprising one reaction step,  
wherein citronellal of general formula (I)

(II)



is cyclized biocatalytically to the corresponding isopulegol of formula (I) by means of an enzyme having the activity of citronellal-isopulegol cyclase.

The above second problem could, surprisingly, be solved by providing mutants of wild-type enzymes, such as Zm-SHC-1 (SEQ ID NO:2). In particular it was in fact found that through targeted introduction of mutations in at least one highly conserved sequence position in said cyclases, in particular squalene-hopene cyclases (cf. alignment of SEQ ID NOs. 2 to 326, below) the enzymatic activity can be influenced in the desired manner.

#### DESCRIPTION OF THE FIGURES

FIG. 1a shows the wild-type amino acid sequence (SEQ ID NO: 2) of squalene-hopene cyclase 1 from *Zymomonas mobilis* (Zm-SHC-1). Position 486 of saturation mutagenesis is marked.

FIG. 1b shows the wild-type nucleic acid sequence (SEQ ID NO: 1) of Zm-SHC-1. Positions 1456-1458 of saturation mutagenesis are marked.

FIG. 2 shows the turnover of the SHC 1 WT protein compared with the F486A mutant as a function of time with 10 mM R(+)- and S(-)-citronellal as substrate. The percentage distribution of substrate and isopulegol product isomers after incubation for various times at 30° C. is shown in each case. Citronellal (diamonds), isopulegol I (squares), isopulegol II (triangles) and isopulegol III (crosses).

FIG. 3 shows the turnover of the various mutants of Zm-SHC-1 compared with the wild type (wt) and the control without enzyme (K) with 10 mM citronellal racemate as substrate. The percentage distribution of substrate and isopulegol product isomers after incubation overnight at 30° C. is shown in each case.

FIG. 4 shows the turnover of the various Zm-SHC-1 mutants compared with the wild type (wt) and the control without enzyme (K) with 25 mM squalene as substrate in the presence of 1% Triton. The percentage distribution of squalene and hopene after incubation for 70 h at 30° C. is shown in each case.

FIGS. 5 to 7 show the reaction of in each case 20 mM substrate after incubation overnight with the mutants Ap-SHC: F481C, Bj-SHC: F447C, Sc-SHC: F449C, Zm SHC-2: F438C and Zm SHC-1 compared with the control; the substrates were citronellal racemate in FIG. 5, R(+)-citronellal in FIG. 6 and S(-)-citronellal in FIG. 7.

#### DETAILED DESCRIPTION OF THE INVENTION

##### A. General Definitions

"Cyclases" in the sense of the present invention are generally enzymes or enzyme mutants, which in particular display the activity of a citronellal-isopulegol cyclase. Intramolecular transferases from the isomerase subclass are suitable as enzymes with the activity of a citronellal-isopulegol cyclase; i.e. proteins with the EC number EC 5.4. (Enzyme code according to Eur. J. Biochem. 1999, 264, 610-650). In particular they are representatives of EC 5.4.99.17. Suitable enzymes with the activity of a citronellal-isopulegol cyclase are in particular those cyclases that also bring about the cyclization of homofarnesol to ambroxan or of squalene to hopene (hence sometimes also designated "SHC": squalene hopene cyclase) and which are described in detail in international application PCT/EP2010/057696, to which reference is expressly made here. In

particular, cyclases according to the invention are those that are derived by mutation of SHCs.

On the basis of the reversibility of enzymatic reactions, the present invention relates to the enzymatic reactions described herein in both directions of reaction.

"Functional mutants" of a "cyclase" include the "functional equivalents" of such enzymes defined below.

The term "biocatalytic process" refers to any process carried out in the presence of catalytic activity of a "cyclase" according to the invention or of an enzyme with "cyclase activity", i.e. processes in the presence of raw, or purified, dissolved, dispersed or immobilized enzyme, or in the presence of whole microbial cells, which have or express such enzyme activity. Biocatalytic processes therefore include both enzymatic and microbial processes.

The term "stereospecific" means that one of several possible stereoisomers of a compound produced according to the invention is produced with at least one asymmetry center by the action of an enzyme according to the invention in high "enantiomeric excess" or high "enantiomeric purity", for example at least 90% ee, in particular at least 95% ee, or at least 98% ee, or at least 99% ee. The ee % value is calculated from the following formula:

$$ee\% = \frac{[X_A - X_B]}{[X_A + X_B]} * 100,$$

in which  $X_A$  and  $X_B$  stand for the mole fraction of enantiomers A and B respectively.

"First sphere residues" and "second sphere residues" are amino acid residues which, based on structural analyses of the protein, are assigned a special proximity to the reactive center of the cyclase. The criterion for the first sphere is the distance from the ligand 2-azasqualene, which is given in a published x-ray structure (pdb: 1 ump). These residues were determined automatically with a computer program (lignin.weizmann.ac.il/cgi-bin/lpccsu/LpcCsu.cgi; Sobolev V, Sorokine A, Prilusky J, Abola E E, Edelman M. Automated analysis of interatomic contacts in proteins. Bioinformatics 1999; 15(4):327-332.). This program assumes that two molecules are in contact with each other when the distance between their atoms corresponds to the sum of their van der Waals radii  $\pm 1$  Å. The second sphere includes all amino acids that are located in a radius of 5 Å to each residue of the first sphere. Such residues therefore appear to be especially suitable for undertaking directed mutation, for further targeted modification of the enzyme activity.

"Cyclase activity", determined with a "reference substrate under standard conditions", is e.g. an enzyme activity that describes the formation of a cyclic product from a noncyclic substrate. Standard conditions are e.g. substrate concentrations from 10 mM to 0.2 M, in particular 15 to 100 mM, for example about 20 to 25 mM; at pH 4 to 8, and at temperatures of e.g. 15 to 30 or 20 to 25° C. It can be determined with recombinant cyclase-expressing cells, lysed cyclase-expressing cells, fractions thereof or enriched or purified cyclase enzyme. In particular the reference substrate is a citronellal of formula (II); in particular R(+)-citronellal, or a citronellal racemate, in a concentration from 15 to 100 mM or about 20 to 25 mM, at 20 to 25° C. and pH 4-6, such as 4.5; as is also described in more detail in the examples.

An "F486-analog" position corresponds to position F486 according to SEQ ID NO:2 from the functional standpoint and can be determined by sequence alignment of SHCs from organisms other than *Zymomonas mobilis* as explained herein. For example the F486-analog position of SEQ ID NO:3 is position F449 and of SEQ ID NO:4 position F481 and of SEQ ID NO:5 position F447 and of SEQ ID NO:6 position F438. Corresponding analogies apply to the other

5

sequence positions described concretely for SEQ ID NO: 2 herein, such as the so-called “first sphere residues” and “second sphere residues” or of the DXDD motif and their analogous positions in SEQ ID NO:3 to 326).

“Terpenes” are hydrocarbons that are made up of isoprene units (C5 units), in particular noncyclic terpenes, for example squalene, the carbon number of which is divisible by 5.

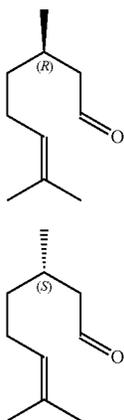
“Terpenoids” are substances that are derived from terpenes, in particular noncyclic terpenes, e.g. by additional insertion of carbon atoms and/or heteroatoms, for example citronellal.

“Terpene-like” compounds for the purposes of the present invention comprise in particular those compounds which fall within the general structural formula (IV) as defined below.

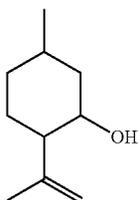
Generally encompassed in accordance with the invention are all isomeric forms of the compounds described herein, such as constitutional isomers and more particularly stereoisomers and mixtures thereof, such as optical isomers or geometric isomers, such as E- and Z-isomers, and also combinations thereof. Where there are two or more centers of asymmetry in a molecule, the invention encompasses all combinations of different conformations of these centers of asymmetry, such as pairs of enantiomers, for example.

“Menthol” encompasses all stereoisomeric forms such as (+)-menthol, (+)-isomenthol, (+)-neomenthol, (+)-neoisomentol, (-)-menthol, (-)-isomenthol, (-)-neomenthol, (-)-neoisomentol and any desired mixtures thereof.

Citronellal of formula (II) is commercially available both as R(+)-citronellal of formula (R-II) and as S(-)-citronellal of formula (S-II) and as racemate of formula (II).



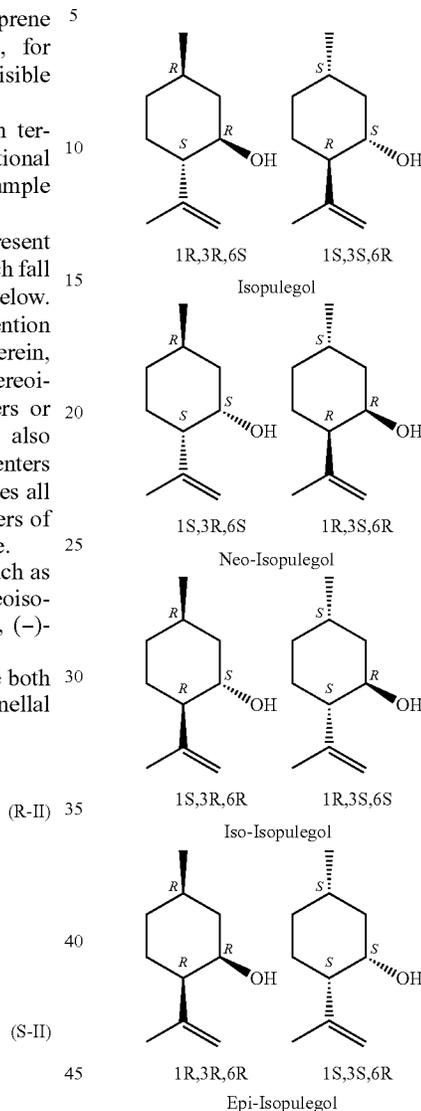
Isopulegol of formula (I)



has in positions 1, 3 and 6 in each case an optically active center, so that in principle 4 different diastereomers with in

6

each case 2 enantiomers, thus altogether 8 stereoisomers, are conceivable, starting from the racemate of citronellal of formula (I).



Isopulegol is also called isopulegol I; neo-isopulegol is also called isopulegol II; iso-isopulegol is also called isopulegol III; epi-isopulegol or neo-iso-isopulegol is also called isopulegol IV.

Unless indicated otherwise, the general chemical definitions that apply herein are as follows:

(I) Alkyl and also all alkyl moieties in radicals derived therefrom, such as hydroxyalkyl, for example: saturated, straight-chain or branched hydrocarbon radicals having 1 to 4, 1 to 6, 1 to 8 or 1 to 10 carbon atoms, e.g.

C<sub>1</sub>-C<sub>6</sub>-alkyl: such as methyl, ethyl, propyl, 1-methylethyl, butyl, 1-methylpropyl, 2-methylpropyl and 1,1-dimethylethyl as exemplary representatives of C<sub>1</sub>-C<sub>4</sub>-alkyl; and also pentyl, 1-methylbutyl, 2-methylbutyl, 3-methylbutyl, 2,2-dimethylpropyl, 1-ethylpropyl, hexyl, 1,1-dimethylpropyl, 1,2-dimethylpropyl, 1-methylpentyl, 2-methylpentyl, 3-methylpentyl, 4-methylpentyl, 1,1-dimethylbutyl, 1,2-dimethylbutyl, 1,3-dimethylbutyl, 2,2-dimethylbutyl, 2,3-dimethylbutyl, 3,3-dimethylbu-

tyl, 1-ethylbutyl, 2-ethylbutyl, 1,1,2-trimethylpropyl, 1,2,2-trimethylpropyl, 1-ethyl-1-methylpropyl and 1-ethyl-2-methylpropyl.

Hydroxy-C<sub>1</sub>-C<sub>6</sub>-alkyl, comprising hydroxy-C<sub>1</sub>-C<sub>4</sub>-alkyl, such as e.g. hydroxymethyl, 1- or 2-hydroxyethyl, 1-, 2- or 3-hydroxypropyl, 1-hydroxymethylethyl, 1-, 2-, 3- or 4-hydroxybutyl, 1-hydroxymethylpropyl and 2-hydroxymethylpropyl.

Alkenyl stands for mono- or polyunsaturated, more particularly monounsaturated, straight-chain or branched hydrocarbon radicals having 2 to 4, 2 to 6, 2 to 8, 2 to 10 or 2 to 20 carbon atoms and one double bond in any desired position, e.g. C<sub>2</sub>-C<sub>6</sub>-alkenyl such as ethenyl, 1-propenyl, 2-propenyl, 1-methylethenyl, 1-butenyl, 2-butenyl, 3-butenyl, 1-methyl-1-propenyl, 2-methyl-1-propenyl, 1-methyl-2-propenyl, 2-methyl-2-propenyl, 1-pentenyl, 2-pentenyl, 3-pentenyl, 4-pentenyl, 1-methyl-1-butenyl, 2-methyl-1-butenyl, 3-methyl-1-butenyl, 1-methyl-2-butenyl, 2-methyl-2-butenyl, 3-methyl-2-butenyl, 1-methyl-3-butenyl, 2-methyl-3-butenyl, 3-methyl-3-butenyl, 1,1-dimethyl-2-propenyl, 1,2-di methyl-1-propenyl, 1,2-dimethyl-2-propenyl, 1-ethyl-1-propenyl, 1-ethyl-2-propenyl, 1-hexenyl, 2-hexenyl, 3-hexenyl, 4-hexenyl, 5-hexenyl, 1-methyl-1-pentenyl, 2-methyl-1-pentenyl, 3-methyl-1-pentenyl, 4-methyl-1-pentenyl, 1-methyl-2-pentenyl, 2-methyl-2-pentenyl, 3-methyl-2-pentenyl, 4-methyl-2-pentenyl, 1-methyl-3-pentenyl, 2-methyl-3-pentenyl, 3-methyl-3-pentenyl, 4-methyl-3-pentenyl, 1-methyl-4-pentenyl, 2-methyl-4-pentenyl, 3-methyl-4-pentenyl, 4-methyl-4-pentenyl, 1,1-dimethyl-2-butenyl, 1,1-dimethyl-3-butenyl, 1,2-dimethyl-1-butenyl, 1,2-dimethyl-2-butenyl, 1,2-dimethyl-3-butenyl, 1,3-dimethyl-1-butenyl, 1,3-dimethyl-2-butenyl, 1,3-dimethyl-3-butenyl, 2,2-dimethyl-3-butenyl, 2,3-dimethyl-1-butenyl, 2,3-dimethyl-2-butenyl, 2,3-dimethyl-3-butenyl, 3,3-dimethyl-1-butenyl, 3,3-dimethyl-2-butenyl, 1-ethyl-1-butenyl, 1-ethyl-2-butenyl, 1-ethyl-3-butenyl, 2-ethyl-1-butenyl, 2-ethyl-2-butenyl, 2-ethyl-3-butenyl, 1,1,2-trimethyl-2-propenyl, 1-ethyl-1-methyl-2-propenyl, 1-ethyl-2-methyl-1-propenyl and 1-ethyl-2-methyl-2-propenyl.

“Oxo”, for example, is a radical which together with the C atom to which it is bonded forms a keto group (C=O).

“Methylene” (=CH<sub>2</sub>), for example, is a radical which together with the C atom to which it is bonded forms a vinyl radical (—CH=CH<sub>2</sub>).

#### B. Special Embodiments of the Invention

The present invention relates in particular to the following special embodiments:

1. Enzyme mutant with cyclase activity, selected from mutants of a wild-type enzyme, which comprises an amino acid sequence, selected from SEQ ID NO: 2 to 326 or a partial sequence thereof; wherein the mutant catalyzes at least the cyclization of at least one citronellal isomer (or a mixture of isomers, for example racemate) according to the above definition to at least one isopulegol isomer (or to a pair of diastereomers I to IV, for example I and/or II) according to the above definition, wherein the partial sequence or short form of the cyclase comprises e.g. at least 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650 or 700 continuous amino acid residues of one of these sequences, and is accessible e.g. by N- and/or C-terminal shortening of the concrete sequences.
2. Enzyme mutant according to embodiment 1, comprising
  - a) a mutation in position F486 of SEQ ID NO: 2 or
  - b) a mutation in a sequence selected from SEQ ID NO: 3 to 326, wherein the mutated position corresponds to position F486 of SEQ ID NO: 2 (i.e. is an “F486-analog” position);

wherein at least the cyclization of at least one citronellal isomer to at least one isopulegol isomer is made possible by the mutation (i.e. the corresponding original or wild-type protein did not catalyze this reaction) or is modified (i.e. the corresponding original or wild-type protein catalyzed this reaction, but e.g. at lower product yield, turnover rate and/or stereospecificity). Moreover, the partial sequence or short form of the cyclase also has this cyclase-typical mutation in a position corresponding to F486 from SEQ ID NO: 2. For example, an N-terminally shortened version of the cyclase according to SEQ ID NO: 2 is an example of said short version. This is characterized by the following N-terminus: (M)  
 KIFGAEKTSYKSPASDTHIGTDTLKRPN . . . wherein the N-terminal K corresponds to position 16 of SEQ ID NO:2.

3. Enzyme mutant according to one of the preceding embodiments in which up to 25% or up to 20, 15, 10, 9, 8, 7, 6, 5, 4, 3, 2 or 1% of the amino acid residues, for example 1 to 30, 2 to 25, 3 to 20 or 4 to 15 or 5 to 10 of the amino acid residues, are in each case altered relative to the unmutated wild-type sequence according to SEQ ID NO: 2 to 326, by deletion, insertion, substitution, addition, inversion or a combination thereof.
4. Enzyme mutant according to one of the preceding embodiments, in which the mutation in position F486 of SEQ ID NO:2 or in a position corresponding to this position in one of the sequences according to SEQ ID NO: 3 to 326, is a substitution selected from F486N, F486Q, F486L, F486M, F486E, F486G, F486S, F486V, F486T, F486C, F486I and F486A or optionally selected from F486H, F486Y, F486W and F486D.
5. Enzyme mutant according to one of the preceding embodiments, in which additionally (or alternatively, but in particular additionally) at least one, for example 1, 2, 3, 4, 5, 6, 7, or 8, mutations in one of the positions W374, D437, D440, F428, W555, Y561, Y702, Y705 (the so-called “first sphere residues”) of SEQ ID NO: 2 or in at least one corresponding position selected from these positions, is present in one of the sequences according to SEQ ID NO: 3 to 326.
6. Enzyme mutant according to one of the preceding embodiments, in which there is no mutation in position D437 and/or D439 and/or D440 of SEQ ID NO: 2 (DXDD motif) or the respective corresponding position in one of the sequences according to SEQ ID NO: 3 to 326.
7. Enzyme mutant according to one of the preceding embodiments, in which there is no mutation in position Y702 of SEQ ID NO: 2 or in the corresponding position in one of the sequences according to SEQ ID NO: 3 to 326, or if a mutation is present, this is a substitution Y702F or optionally Y702E or Y702D or corresponding substitution.
8. Enzyme mutant according to one of the preceding embodiments, which optionally is further mutated in at least one, for example 1 to 15, 1 to 10 or 1 to 5, such as 1, 2, 3 or 4, of positions P229, D439, D508, E601, G553, G556, N432, P436, P499, R224, S371, T376, T563, W414 or W624 (the so-called “second sphere residues”) of SEQ ID NO: 2 or in at least one corresponding position selected from these positions, in one of the sequences according to SEQ ID NO: 3 to 326; and optionally a further mutation in position E429, L700 and R554 of SEQ ID NO: 2 or the analogous positions of SEQ ID NO: 3 to 326.

9

9. Enzyme mutant according to one of the preceding embodiments, selected from

a) the single mutants

F486X with X=N, Q, L, M, E, G, S, V, T, C, I or A according to SEQ ID NO: 2 or a short version thereof;

Y702X with X=F, A, C or S according to SEQ ID NO: 2 or a short version thereof;

Y561X with X=A or S according to SEQ ID NO: 2 or a short version thereof;

wherein the short version comprises e.g. the following N-terminal sequence:

(M) KIFGAEKTSYKPASDTIIGTDTLKRPN . . .

b) the multiple mutants F486A/Y702A, F486A/Y561A or F486A/Y705A according to SEQ ID NO: 2

c) the mutants corresponding to a) or b), derived from one of SEQ ID NO: 3 to 325.

10. Enzyme mutant according to one of the preceding embodiments, which comprises at least 50%, for example 50 to 100% or more than 100%, for example >100 to 1000%, in each case determined under standard conditions using a reference substrate that displays citronellal-isopulegol cyclase activity of an enzyme, which has an amino acid sequence according to SEQ ID NO: 2 from position 1 to 725, 2 to 725 or 16 to 725, optionally extended N-terminally with a methionine residue.

11. Enzyme mutant according to embodiment 10, wherein the citronellal-isopulegol cyclase activity is determined under standard conditions using a citronellal, for example the racemate or the R(+) form, as reference substrate.

12. Enzyme mutant according to one of the preceding embodiments, wherein the mutation takes place in an enzyme, and comprises an amino acid sequence according to SEQ ID NO: 2 from position 1 to 725, 2 to 725 or 16 to 725, optionally extended N-terminally with a methionine residue.

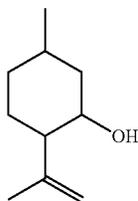
13. Nucleic acid sequence coding for a mutant according to one of the preceding embodiments.

14. Expression cassette, comprising a nucleic acid sequence according to embodiment 13.

15. Recombinant vector, comprising, under the control of at least one regulatory element, at least one nucleic acid sequence according to embodiment 13 or at least one expression cassette according to embodiment 14.

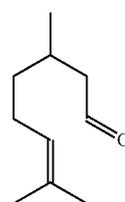
16. Recombinant microorganism, comprising at least one nucleic acid sequence according to embodiment 13 or at least one expression cassette according to embodiment 14 or at least one vector according to embodiment 15.

17. Biocatalytic process for producing isopulegol of general formula (I)



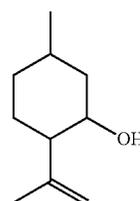
10

wherein citronellal of general formula (II)

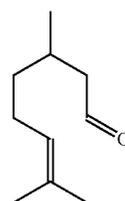


is cyclized to isopulegol of formula (I) by means of an enzyme of EC class EC 5.4.99, in particular of EC class EC 5.4.99.17, or in the presence of a microorganism expressing this enzyme.

18. Biocatalytic process for producing isopulegol of general formula (I)

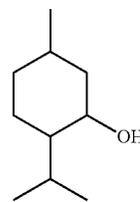


wherein citronellal of general formula (II)



is cyclized to isopulegol of formula (I) by means of an enzyme mutant according to one of embodiments 1 to 12, or in the presence of a microorganism expressing this enzyme mutant according to embodiment 16.

19. A method of production of menthol of formula III



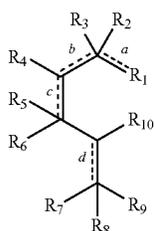
by  
a) cyclizing citronellal to isopulegol by a method according to embodiment 17 or 18, and  
b) catalytically hydrogenating isopulegol to menthol.

20. The method according to embodiment 19, where the hydrogenation takes place in the presence of hydrogen and a catalyst comprising

## 11

30% to 70% by weight of oxygen-containing compounds of nickel, calculated as NiO,  
 15% to 45% by weight of oxygen-containing compounds of zirconium, calculated as ZrO<sub>2</sub>,  
 5% to 30% by weight of oxygen-containing compounds of copper, calculated as CuO, and  
 0.1% to 10% by weight of oxygen-containing compounds of molybdenum, calculated as MoO<sub>3</sub>,  
 the % by weight figures being based on the dry, unreduced catalyst.

21. A method for enzymatic or biocatalytic conversions of compounds of general formula IV



in which

“a”, “b”, “c” and “d”, in each case independently of one another, represent a single or double C—C bond, with the proviso that cumulative double bonds are excluded; and with the following provisos:

R<sub>1</sub> possesses the following definitions:

(1) when “a” is a double bond:

R<sub>1</sub> is selected from  
 oxo (=O), or  
 CH—(CH<sub>2</sub>)<sub>n</sub>—Z,

in which n is 0, 1 or 2 and

Z is OH, CHO, C(O)alkyl, such as C(O)C<sub>1</sub>-C<sub>4</sub>-alkyl, in particular C(O)—CH<sub>3</sub> or C(O)—CH<sub>2</sub>CH<sub>3</sub>; COOH, C(CH<sub>2</sub>)—CH=CH<sub>2</sub>; C(OH)(CH<sub>3</sub>)—CH=CH<sub>2</sub>; C(CH<sub>3</sub>)=CH—CH=CH<sub>2</sub>; or a radical of the formula C(CH<sub>3</sub>)=CH—CH<sub>2</sub>Y

in which

Y is OH, CH<sub>2</sub>OH, COOH, or CH<sub>2</sub>C(O)CH<sub>3</sub>; or

(2) when “a” is a single bond:

R<sub>1</sub> is selected from  
 CH<sub>3</sub>; CHO; CH<sub>2</sub>CH<sub>2</sub>OH; CH=CH<sub>2</sub>; CH<sub>2</sub>C(O)OH; CH<sub>2</sub>CHO or C<sub>3</sub>H<sub>6</sub>CH(CH<sub>3</sub>)CHO;

wherein, when “a” is a double bond, it has E or Z configuration;

R<sub>2</sub> and R<sub>3</sub> possess the following definitions:

(1) when “a” and “b” are each a single bond:

R<sub>2</sub> and R<sub>3</sub> independently of one another are H, alkyl, such as C<sub>1</sub>-C<sub>4</sub>-alkyl or OH, or R<sub>2</sub> and R<sub>3</sub> together are a methylene (=CH<sub>2</sub>) or oxo (=O) group; or

(2) when “a” or “b” is a double bond, one of the radicals R<sub>2</sub> and R<sub>3</sub> is absent and the other of the two radicals is H, C<sub>1</sub>-C<sub>4</sub>-alkyl, in particular methyl, or OH;

R<sub>4</sub> is H or hydroxy-C<sub>1</sub>-C<sub>4</sub>-alkyl, in particular Hydroxymethyl;

R<sub>5</sub> and R<sub>6</sub> possess the following definitions:

(1) when “c” is a single bond:

R<sub>5</sub> and R<sub>6</sub> are each H, or R<sub>5</sub> and R<sub>6</sub> together are an oxo (=O) group; or

(2) when “c” is a double bond, one of the radicals R<sub>5</sub> and R<sub>6</sub> is absent and the other of the two radicals is H;

## 12

R<sub>7</sub>, R<sub>8</sub> and R<sub>9</sub> possess the following definitions:

(1) when “d” is a single bond:

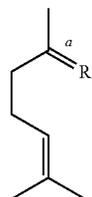
two of the radicals R<sub>7</sub>, R<sub>8</sub> and R<sub>9</sub> in each case independently of one another are H or alkyl, such as C<sub>1</sub>-C<sub>4</sub>-alkyl, in particular methyl or ethyl, and the other of the radicals is OH; or

(2) when “d” is a double bond, one of the radicals R<sub>7</sub>, R<sub>8</sub> and R<sub>9</sub> is absent and the other of the two radicals in each case independently of one another are H or alkyl, such as C<sub>1</sub>-C<sub>4</sub>-alkyl, in particular methyl or ethyl;

R<sub>10</sub> is H or hydroxy-C<sub>1</sub>-C<sub>6</sub>-alkyl, such as hydroxy-C<sub>1</sub>-C<sub>4</sub>-alkyl, or mono- or polyunsaturated C<sub>2</sub>-C<sub>6</sub>-alkenyl, such as, in particular, H or CH=CH—C(CH<sub>3</sub>)=CH<sub>2</sub>;

where a compound of the formula IV in stereoisomerically pure form, or a stereoisomer mixture thereof, is reacted using an enzyme of class EC 5.4.99, in particular of class EC 5.4.99.17, or an enzyme mutant according to one of embodiments 1 to 12 or in the presence of a microorganism according to embodiment 16 expressing these enzymes or enzyme mutants.

22. The method according to embodiment 21, in which a compound is converted which is selected from compounds of the formula IVa



in which R<sub>1</sub> possesses the definitions indicated above and in particular is the radical CH—(CH<sub>2</sub>)<sub>n</sub>—Z

in which

n=0 and Z=CHO, or COOK or

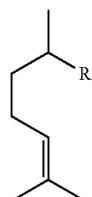
n=1 and Z=OH; or

n=2 and Z=C(O)CH<sub>3</sub>; COOH, C(CH<sub>2</sub>)—CH=CH<sub>2</sub>; C(CH<sub>3</sub>)=CH—CH=CH<sub>2</sub>;

or is a radical of the formula C(CH<sub>3</sub>)=CH—CH<sub>2</sub>Y in which Y is OH, CH<sub>2</sub>OH, COOH, or CH<sub>2</sub>C(O)CH<sub>3</sub>;

and “a” optionally has E or Z configuration;

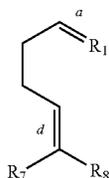
or of the formula IVb



in which R<sub>1</sub> possesses the definitions indicated above and in particular is CH<sub>2</sub>CHO;

13

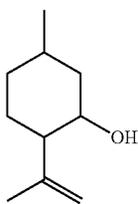
or of the formula IVc



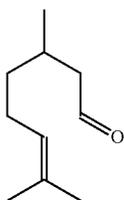
in which

R<sub>1</sub> possesses the definitions indicated above, and in particular is CH—CHO; and one of the radicals R<sub>7</sub> and R<sub>8</sub> is H and the other is C<sub>1</sub>-C<sub>4</sub>-alkyl, where in particular R<sub>7</sub> is ethyl and the double bonds “a” and “d” have Z configuration.

23. The method according to one of embodiments 20 to 22, in which the compound of the formula IV is selected from citronellal; citral; farnesol; homofarnesol; homofarnesol derivatives, such as homofarnesyllic acid; geranylacetone, melonal; nonadienal; and trimethyldecatetraene.
24. Use of an enzyme from EC class EC 5.4.99, in particular from EC class EC 5.4.99.17 for the cyclization of terpenes and/or terpenoids, in particular for the conversion of citronellal to isopulegol.
25. Use of an enzyme mutant according to one of embodiments 1 to 12, a nucleic acid according to embodiment 13, an expression construct according to embodiment 14, a recombinant vector according to embodiment 15 or a recombinant microorganism according to embodiment 1 for the cyclization of terpenes and/or terpenoids, and for the conversion of compounds of the general formula IV according to the definition in one of the embodiments 20 to 23.
25. Use according to embodiment 25 for the conversion of citronellal to isopulegol; or for the conversion of squalene to hopene.
26. A method of production of isopulegol of general formula (I)



comprising one reaction step,  
wherein citronellal of general formula (II)



14

is cyclized biocatalytically to the corresponding isopulegol of formula (I) by means of an enzyme having the activity of a citronellal-isopulegol cyclase.

- (IVc) 27. The method according to embodiment 26, wherein the enzyme possesses a polypeptide sequence which either
- 5 a) is SEQ ID NO: 2, or
- b) in which up to 25%, such as, for example, up to 20, 15, 10, 9, 8, 7, 6, 5, 4, 3, 2 or 1% of the amino acid residues are altered relative to SEQ ID NO: 2 by deletion, insertion, substitution or a combination thereof, and
- 10 which still has at least 50%, such as, for example, at least 60, 65, 70, 75, 80, 85, 90 or 95%, of the enzymatic activity of SEQ ID NO: 2.
28. The method according to embodiment 26 or 27, wherein the enzyme is encoded by a nucleic acid sequence according to SEQ ID NO: 1 or a functional equivalent thereof.
29. The method according to one of embodiments 26 to 28, wherein the enzyme is encoded by a nucleic acid sequence according to SEQ ID NO: 1 or a functional equivalent thereof, the nucleic acid sequence being part of a gene construct or vector.
30. The method according to one of embodiments 26 to 29, wherein the enzyme is encoded by a nucleic acid sequence according to SEQ ID NO: 1 or a functional equivalent thereof, the nucleic acid sequence being part of a gene construct or vector which are present in a host cell.
31. The method according to one of embodiments 26 to 30, wherein the enzyme is present in a form selected from the group consisting of:
- a) free, optionally purified or partly purified polypeptide having the activity of a citronellal-isopulegol cyclase;
- b) immobilized polypeptide having the activity of a citronellal-isopulegol cyclase;
- c) polypeptide according to a) or b) which is isolated from cells;
- d) whole cell, optionally resting or digested cells, comprising at least one polypeptide having the activity of a citronellal-isopulegol cyclase;
- e) cell lysate or cell homogenate of the cells described under d).
32. The method according to embodiment 31, wherein the cells are microorganisms, preferably transgenic microorganisms expressing at least one heterologous nucleic acid molecule coding for a polypeptide having the activity of a citronellal-isopulegol cyclase.
- (I) 33. The method according to one of embodiments 26 to 32, wherein the production of isopulegol takes place in one-phase aqueous systems or in two-phase systems.
34. The method according to one of embodiments 26 to 33, in which the reaction of citronellal to isopulegol takes place at a temperature in the range from 20 to 40° C. and/or at a pH in the range from 4 to 8.
35. The method according to one of embodiments 26 to 34, wherein the enzyme having the activity of a citronellal-isopulegol cyclase is encoded by a gene which has been isolated from a microorganism selected from the group of microorganisms consisting of *Zymomonas mobilis*, *Methylococcus capsulatus*, *Rhodospseudomonas palustris*, *Bradyrhizobium japonicum*, *Frankia spec.* and *Streptomyces coelicolor*, in particular *Zymomonas mobilis*.
- (II) 36. The method according to one of embodiments 26 to 35, wherein the enzyme having the activity of a citronellal-isopulegol cyclase has been produced by a microorganism which overproduces the enzyme having the activity of a citronellal-isopulegol cyclase and which has been selected from the group of microorganisms consisting of the genera *Escherichia*, *Corynebacterium*, *Ralstonia*,
- 65

*Clostridium*, *Pseudomonas*, *Bacillus*, *Zymomonas*, *Rhodobacter*, *Streptomyces*, *Burkholderia*, *Lactobacillus* and *Lactococcus*.

37. The method according to one of embodiments 26 to 36, wherein the enzyme having the activity of a citronellal-isopulegol cyclase has been produced by transgenic microorganisms of the species *Escherichia coli*, *Pseudomonas putida*, *Burkholderia glumae*, *Corynebacterium glutamicum*, *Saccharomyces cerevisiae*, *Pichia pastoris*, *Streptomyces lividans*, *Streptomyces coelicolor*, *Bacillus subtilis* or *Zymomonas mobilis* which overproduce the enzyme having the activity of a citronellal-isopulegol cyclase.
38. Use of an enzyme having the activity of a citronellal-isopulegol cyclase for the biocatalytic conversion of citronellal to isopulegol.
39. Use according to embodiment 38, wherein the enzyme possesses a polypeptide sequence which either
- is SEQ ID NO: 2, or
  - in which up to 25%, such as, for example, up to 20, 15, 10, 9, 8, 7, 6, 5, 4, 3, 2 or 1% of the amino acid residues are altered relative to SEQ ID NO: 2 by deletion, insertion, substitution or a combination thereof, and which still has at least 50%, such as, for example, at least 60, 65, 70, 75, 80, 85, 90 or 95%, of the enzymatic activity of SEQ ID NO: 2.

40. Use according to embodiment 38 or 39, wherein the enzyme is encoded by a nucleic acid sequence according to SEQ ID NO: 1 or a functional equivalent thereof.
41. Use of a gene construct or vector comprising a nucleic acid sequence according to SEQ ID NO: 1 or a functional equivalent thereof, which encode a polypeptide having the activity of a citronellal-isopulegol cyclase, which serves for the biocatalytic conversion of citronellal to isopulegol, in a method of production of isopulegol by cyclization of citronellal.
42. Use of a host cell which comprises a gene construct or a vector comprising a nucleic acid sequence according to SEQ ID NO: 1 or a functional equivalent thereof, for preparing an enzyme having the activity of a citronellal-isopulegol cyclase for the biocatalytic conversion of citronellal to isopulegol.

### C. Further Embodiments of the Invention

#### 1. Especially Suitable Wild-type Sequences

SHC wild-type sequences usable according to the invention, whose SEQ ID NO, source organism, GenBank reference number, the amino acid residue "corresponding" to position F486 of SEQ ID NO:2, i.e. F486-analog ("Aa") and whose sequence position are presented in the following table. The information is based on a sequence alignment, which was set up as follows:

---

Program: CLUSTALW,  
Default parameters:  
Protein Gap Open Penalty 10.0  
Protein Gap Extension Penalty 0.2  
Protein weight matrix: Gonnet series

---

S_ID DB	SEQ ID NO	Organism	GI No. of the reference sequences	Aa	Position
s1	seq_ID 2	<i>Zymomonas mobilis</i>	AAV90172.1	F	486
s20	seq_ID 3	<i>Streptomyces coelicolor</i>	CAB39697.1	F	449
s911	seq_ID 4	<i>Acetobacter pasteurianus</i>	BAH99456.1	F	481
s2	seq_ID 5	<i>Bradyrhizobium</i> sp.	ABQ33590.1	F	447
s940	seq_ID 6	<i>Zymomonas mobilis</i>	EER62728.1	F	438
s949	seq_ID 7	<i>Acidithiobacillus caldus</i>	EET25937.1	Y	432
s167	seq_ID 8	<i>Acidithiobacillus ferrooxidans</i>	ACH84004.1	Y	429
s41	seq_ID 9	<i>Acidobacterium capsulatum</i>	ACO34244.1	F	458
s36	seq_ID 10	<i>Acidothermus cellulolyticus</i>	ABK53469.1	F	426
s83	seq_ID 11	<i>Adiantum capillus-veneris</i>	BAF93209.1	Y	436
s143	seq_ID 12	<i>Ajellomyces capsulatus</i>	EDN09769.1	F	496
s995	seq_ID 13	<i>Ajellomyces capsulatus</i>	EER40510.1	—	432
s163	seq_ID 14	<i>Ajellomyces capsulatus</i>	EEH02950.1	F	429
s13	seq_ID 15	<i>Alicyclobacillus acidocaldarius</i>	EED08231.1	Y	420
s14	seq_ID 16	<i>Alicyclobacillus acidocaldarius</i>	P33247.4	Y	420
s1193	seq_ID 17	<i>Alicyclobacillus acidocaldarius</i>	AAT70690.1	Y	116
s21	seq_ID 18	<i>Alicyclobacillus acidoterrestris</i>	CAA61950.1	Y	420
s1189	seq_ID 19	<i>Alicyclobacillus acidoterrestris</i>	AAT70691.1	Y	121
s51	seq_ID 20	<i>Anabaena variabilis</i>	ABA24268.1	F	423
s76	seq_ID 21	<i>Anaeromyxobacter</i> sp.	ABS28257.1	F	440
s159	seq_ID 22	<i>Aspergillus clavatus</i>	EAW07713.1	F	446
s131	seq_ID 23	<i>Aspergillus flavus</i>	EED48353.1	F	444
s176	seq_ID 24	<i>Aspergillus fumigatus</i>	EDP50814.1	F	502
s126	seq_ID 25	<i>Aspergillus fumigatus</i>	EAL84865.1	F	449
s178	seq_ID 26	<i>Aspergillus fumigatus</i>	EAL86291.2	F	406
s121	seq_ID 27	<i>Aspergillus niger</i>	CAK43501.1	F	441
s115	seq_ID 28	<i>Aspergillus niger</i>	CAK45506.1	F	440
s124	seq_ID 29	<i>Aspergillus oryzae</i>	BAE63941.1	F	444
s119	seq_ID 30	<i>Azotobacter vinelandii</i>	EAM07611.1	F	442
s223	seq_ID 31	<i>Bacillus amyloliquefaciens</i>	ABS74269.1	F	413
s221	seq_ID 32	<i>Bacillus anthracis</i>	AAP27368.1	F	409
s976	seq_ID 33	<i>Bacillus cereus</i>	EEK66523.1	F	423
s225	seq_ID 34	<i>Bacillus cereus</i>	EAL12758.1	F	423
s972	seq_ID 35	<i>Bacillus cereus</i>	EEL44583.1	F	412
s977	seq_ID 36	<i>Bacillus cereus</i>	EEK43841.1	F	412
s985	seq_ID 37	<i>Bacillus cereus</i>	EEK82938.1	F	412
s988	seq_ID 38	<i>Bacillus cereus</i>	EEK99528.1	F	412
s981	seq_ID 39	<i>Bacillus cereus</i>	EEK77935.1	F	412

Program: CLUSTALW,  
 Default parameters:  
 Protein Gap Open Penalty 10.0  
 Protein Gap Extension Penalty 0.2  
 Protein weight matrix: Gonnet series

S_ID DB	SEQ ID NO	Organism	GI No. of the reference sequences	Aa	Position
s987	seq_ID 40	<i>Bacillus cereus</i>	EEL81079.1	F	412
s960	seq_ID 41	<i>Bacillus cereus</i>	EEL88307.1	F	412
s979	seq_ID 42	<i>Bacillus cereus</i>	EEL63943.1	F	412
s974	seq_ID 43	<i>Bacillus cereus</i>	EEL59884.1	F	412
s956	seq_ID 44	<i>Bacillus cereus</i>	EEL69857.1	F	412
s951	seq_ID 45	<i>Bacillus cereus</i>	EEL92663.1	F	412
s986	seq_ID 46	<i>Bacillus cereus</i>	EEL49968.1	F	411
s227	seq_ID 47	<i>Bacillus cereus</i>	AAU16998.1	F	409
s224	seq_ID 48	<i>Bacillus cereus</i>	AAS42477.1	F	409
s212	seq_ID 49	<i>Bacillus cereus</i>	ACK95843.1	F	409
s289	seq_ID 50	<i>Bacillus coahuilensis</i>	205373680	F	276
s219	seq_ID 51	<i>Bacillus cytotoxicus</i>	ABS22481.1	F	411
s230	seq_ID 52	<i>Bacillus licheniformis</i>	AAU23777.1	F	414
s955	seq_ID 53	<i>Bacillus mycoides</i>	EEL98438.1	F	412
s990	seq_ID 54	<i>Bacillus mycoides</i>	EEM04821.1	F	411
s989	seq_ID 55	<i>Bacillus pseudomycooides</i>	EEM16144.1	F	411
s247	seq_ID 56	<i>Bacillus pumilus</i>	ABV62529.1	F	409
s250	seq_ID 57	<i>Bacillus pumilus</i>	EDW21137.1	F	409
s249	seq_ID 58	<i>Bacillus</i> sp.	EAR64404.1	F	425
s218	seq_ID 59	<i>Bacillus</i> sp.	EDL66148.1	F	412
s241	seq_ID 60	<i>Bacillus subtilis</i>	Q796C3.1	F	415
s284	seq_ID 61	<i>Bacillus subtilis</i>	AAB84441.1	F	415
s215	seq_ID 62	<i>Bacillus thuringiensis</i>	ABK86448.1	F	423
s984	seq_ID 63	<i>Bacillus thuringiensis</i>	EEM21409.1	F	412
s957	seq_ID 64	<i>Bacillus thuringiensis</i>	EEM82653.1	F	412
s980	seq_ID 65	<i>Bacillus thuringiensis</i>	EEM52372.1	F	412
s961	seq_ID 66	<i>Bacillus thuringiensis</i>	EEM27851.1	F	412
s969	seq_ID 67	<i>Bacillus thuringiensis</i>	EEM40716.1	F	412
s959	seq_ID 68	<i>Bacillus thuringiensis</i>	EEM46814.1	F	409
s965	seq_ID 69	<i>Bacillus thuringiensis</i>	EEM94969.1	F	409
s202	seq_ID 70	<i>Bacillus weihenstephanensis</i>	ABY44436.1	F	409
s63	seq_ID 71	Bacterium Ellin514	EEF57225.1	F	461
s72	seq_ID 72	Bacterium Ellin514	EEF59508.1	Y	435
s87	seq_ID 73	<i>Beijerinckia indica</i>	ACB96717.1	F	441
s69	seq_ID 74	<i>Blastopirellula marina</i>	EAQ81955.1	F	475
s543	seq_ID 75	<i>Blastopirellula marina</i>	EAQ78122.1	F	389
s156	seq_ID 76	<i>Bradyrhizobium japonicum</i>	CAA60250.1	F	439
s938	seq_ID 77	<i>Acetobacter pasteurianus</i>	BAH98349.1	F	437
s3	seq_ID 78	<i>Bradyrhizobium</i> sp.	CAL79893.1	F	447
s201	seq_ID 79	<i>Brevibacillus brevis</i>	BAH44778.1	F	448
s148	seq_ID 80	<i>Burkholderia ambifaria</i>	EDT05097.1	F	450
s158	seq_ID 81	<i>Burkholderia ambifaria</i>	EDT37649.1	F	450
s149	seq_ID 82	<i>Burkholderia ambifaria</i>	ACB68303.1	F	446
s100	seq_ID 83	<i>Burkholderia ambifaria</i>	EDT42454.1	F	436
s146	seq_ID 84	<i>Burkholderia cenocepacia</i>	EAY66961.1	F	451
s139	seq_ID 85	<i>Burkholderia cenocepacia</i>	ACA95661.1	F	451
s147	seq_ID 86	<i>Burkholderia cenocepacia</i>	CAR57099.1	F	451
s95	seq_ID 87	<i>Burkholderia cenocepacia</i>	CAR56694.1	F	436
s102	seq_ID 88	<i>Burkholderia dolosa</i>	EAY71311.1	F	437
s941	seq_ID 89	<i>Burkholderia glumae</i>	ACR32572.1	F	555
s945	seq_ID 90	<i>Burkholderia glumae</i>	ACR30752.1	F	449
s132	seq_ID 91	<i>Burkholderia graminis</i>	EDT12320.1	F	462
s104	seq_ID 92	<i>Burkholderia mallei</i>	ABM48844.1	F	436
s140	seq_ID 93	<i>Burkholderia multivorans</i>	ABX19650.1	F	450
s116	seq_ID 94	<i>Burkholderia multivorans</i>	ABX16859.1	F	436
s91	seq_ID 95	<i>Burkholderia oklahomensis</i>	167567074	F	447
s111	seq_ID 96	<i>Burkholderia phymatum</i>	ACC73258.1	F	456
s127	seq_ID 97	<i>Burkholderia phytofirmans</i>	ACD21317.1	F	455
s120	seq_ID 98	<i>Burkholderia pseudomallei</i>	EEC32728.1	F	436
s137	seq_ID 99	<i>Burkholderia</i> sp.	EEA03553.1	F	460
s144	seq_ID 100	<i>Burkholderia</i> sp.	ABB06563.1	F	450
s98	seq_ID 101	<i>Burkholderia</i> sp.	ABB10136.1	F	436
s944	seq_ID 102	<i>Burkholderia</i> sp. CCGE1002	EFA54357.1	F	473
s89	seq_ID 103	<i>Burkholderia thailandensis</i>	167840988	F	451
s113	seq_ID 104	<i>Burkholderia thailandensis</i>	167617352	F	442
s154	seq_ID 105	<i>Burkholderia ubonensis</i>	167589807	F	445
s93	seq_ID 106	<i>Burkholderia ubonensis</i>	167584986	F	436
s96	seq_ID 107	<i>Burkholderia vietnamiensis</i>	ABO56791.1	F	436
s150	seq_ID 108	<i>Burkholderia xenovorans</i>	ABE35912.1	F	457

---

Program: CLUSTALW,  
 Default parameters:  
 Protein Gap Open Penalty 10.0  
 Protein Gap Extension Penalty 0.2  
 Protein weight matrix: Gonnet series

---

S_ID DB	SEQ ID NO	Organism	GI No. of the reference sequences	Aa	Position
s54	seq_ID 109	<i>Candidatus Koribacter</i>	ABF40741.1	F	435
s171	seq_ID 110	<i>Candidatus Kuenenia</i>	CAJ71215.1	F	273
s79	seq_ID 111	<i>Candidatus Solibacter</i>	ABJ82180.1	F	439
s99	seq_ID 112	<i>Candidatus Solibacter</i>	ABJ82254.1	F	429
s917	seq_ID 113	<i>Catenulispora acidiphila</i>	ACU75510.1	F	418
s65	seq_ID 114	<i>Chthoniobacter flavus</i>	EDY15838.1	F	433
s637	seq_ID 115	<i>Chthoniobacter flavus</i>	EDY22035.1	F	384
s38	seq_ID 116	<i>Crocospaera watsonii</i>	EAM53094.1	F	426
s186	seq_ID 117	<i>Cupriavidus taiwanensis</i>	CAQ72562.1	F	454
s32	seq_ID 118	<i>Cyanothece</i> sp.	ACB53858.1	F	441
s40	seq_ID 119	<i>Cyanothece</i> sp.	ACK71719.1	F	430
s30	seq_ID 120	<i>Cyanothece</i> sp.	EDY02410.1	F	429
s29	seq_ID 121	<i>Cyanothece</i> sp.	ACK66841.1	F	429
s47	seq_ID 122	<i>Cyanothece</i> sp.	EDX97382.1	F	428
s35	seq_ID 123	<i>Cyanothece</i> sp.	EAZ91809.1	F	426
s39	seq_ID 124	<i>Cyanothece</i> sp.	ACL45896.1	F	423
s925	seq_ID 125	<i>Cyanothece</i> sp. PCC 8802	ACV02092.1	F	429
s64	seq_ID 126	<i>Desulfovibrio salexigens</i>	EEC62384.1	F	475
s74	seq_ID 127	<i>Dryopteris crassirhizoma</i>	BAG68223.1	F	444
s59	seq_ID 128	<i>Frankia alni</i>	CAJ61140.1	Y	533
s48	seq_ID 129	<i>Frankia alni</i>	CAJ60090.1	F	493
s56	seq_ID 130	<i>Frankia</i> sp.	ABD10207.1	F	530
s60	seq_ID 131	<i>Frankia</i> sp.	ABW15063.1	F	512
s31	seq_ID 132	<i>Frankia</i> sp.	ABW14125.1	Y	481
s948	seq_ID 133	<i>Frankia</i> sp. Eul1c	EFA59873.1	F	557
s919	seq_ID 134	<i>Frankia</i> sp. Eul1c	EFA59089.1	F	553
s628	seq_ID 135	<i>Gemmata obscuriglobus</i>	168700710	F	387
s209	seq_ID 136	<i>Geobacillus</i> sp.	EED61885.1	F	404
s206	seq_ID 137	<i>Geobacillus</i> sp.	EDY05760.1	F	403
s964	seq_ID 138	<i>Geobacillus</i> sp. Y412MC52	EEN95021.1	F	404
s993	seq_ID 139	<i>Geobacillus</i> sp. Y412MC61	ACX79399.1	F	404
s205	seq_ID 140	<i>Geobacillus thermodenitrificans</i>	ABO67242.1	F	403
s15	seq_ID 141	<i>Geobacter bemidjiensis</i>	ACH40355.1	F	468
s8	seq_ID 142	<i>Geobacter lovleyi</i>	ACD95949.1	F	470
s62	seq_ID 143	<i>Geobacter metallireducens</i>	ABB30662.1	F	493
s12	seq_ID 144	<i>Geobacter metallireducens</i>	ABB33038.1	F	467
s73	seq_ID 145	<i>Geobacter</i> sp.	ACM21577.1	F	487
s10	seq_ID 146	<i>Geobacter</i> sp.	EDV72707.1	F	468
s11	seq_ID 147	<i>Geobacter</i> sp.	ACM22003.1	F	467
s913	seq_ID 148	<i>Geobacter</i> sp. M18	EET34621.1	F	468
s914	seq_ID 149	<i>Geobacter</i> sp. M21	ACT16952.1	F	468
s58	seq_ID 150	<i>Geobacter sulfurreducens</i>	AAR36453.1	F	493
s7	seq_ID 151	<i>Geobacter sulfurreducens</i>	AAR34018.1	F	467
s9	seq_ID 152	<i>Geobacter urarii</i> <i>reducens</i>	ABQ25226.1	F	467
s46	seq_ID 153	<i>Gloeobacter violaceus</i>	BAC91998.1	F	425
s67	seq_ID 154	<i>Gluconacetobacter diazotrophicus</i>	ACI51585.1	F	444
s165	seq_ID 155	<i>Gluconacetobacter diazotrophicus</i>	CAP55563.1	F	444
s68	seq_ID 156	<i>Gluconobacter oxydans</i>	AAW61994.1	F	445
s80	seq_ID 157	<i>Granulibacter bethesdensis</i>	ABI63005.1	F	429
s937	seq_ID 158	<i>Hyphomicrobium denitrificans</i>	EET65847.1	F	444
s932	seq_ID 159	<i>Leptospirillum ferro-diazotrophum</i>	EES53667.1	F	460
s24	seq_ID 160	<i>Leptospirillum rubarum</i>	EAY57382.1	F	448
s25	seq_ID 161	<i>Leptospirillum</i> sp.	EDZ38599.1	F	448
s174	seq_ID 162	<i>Magnaporthe grisea</i>	EDK02551.1	F	445
s153	seq_ID 163	<i>Magnetospirillum magnetotacticum</i>	46203107	F	447
s49	seq_ID 164	<i>Methyloacidiphilum inferorum</i>	ACD82457.1	F	456
s169	seq_ID 165	<i>Methylobacterium chloromethanicum</i>	ACK83067.1	F	447
s75	seq_ID 166	<i>Methylobacterium chloromethanicum</i>	ACK86232.1	F	426
s946	seq_ID 167	<i>Methylobacterium extorquens</i>	CAX24364.1	F	447
s141	seq_ID 168	<i>Methylobacterium nodulans</i>	ACL61886.1	F	442
s152	seq_ID 169	<i>Methylobacterium populi</i>	ACB79998.1	F	447
s162	seq_ID 170	<i>Methylobacterium radiotolerans</i>	ACB27373.1	F	445
s180	seq_ID 171	<i>Methylobacterium</i> sp.	ACA20611.1	F	442
s175	seq_ID 172	<i>Methylocella silvestris</i>	ACK52150.1	F	451
s181	seq_ID 173	<i>Methylococcus capsulatus</i>	CAA71098.1	F	439

---

Program: CLUSTALW,  
 Default parameters:  
 Protein Gap Open Penalty 10.0  
 Protein Gap Extension Penalty 0.2  
 Protein weight matrix: Gonnet series

---

S_ID DB	SEQ ID NO	Organism	GI No. of the reference sequences	Aa	Position
s55	seq_ID 174	<i>Microcystis aeruginosa</i>	CAO86472.1	F	423
s101	seq_ID 175	<i>Neosartorya fischeri</i>	EAW20752.1	F	448
s129	seq_ID 176	<i>Nitrobacter hamburgensis</i>	ABE63461.1	F	433
s161	seq_ID 177	<i>Nitrobacter</i> sp.	EAQ34404.1	F	430
s160	seq_ID 178	<i>Nitrobacter winogradskyi</i>	ABA05523.1	F	433
s157	seq_ID 179	<i>Nitrococcus mobilis</i>	EAR22397.1	F	436
s164	seq_ID 180	<i>Nitrosococcus oceani</i>	ABA57818.1	F	446
s170	seq_ID 181	<i>Nitrosomonas europaea</i>	CAD85079.1	F	452
s173	seq_ID 182	<i>Nitrosomonas eutropha</i>	ABI59752.1	F	456
s943	seq_ID 183	<i>Nitrosomonas</i> sp. AL212	EET32702.1	F	452
s142	seq_ID 184	<i>Nitrosospora multififormis</i>	ABB75845.1	F	439
s52	seq_ID 185	<i>Nostoc punctiforme</i>	ACC84529.1	F	423
s45	seq_ID 186	<i>Nostoc</i> sp.	BAB72732.1	F	423
s122	seq_ID 187	<i>Oligotropha carboxidovorans</i>	ACI93782.1	F	433
s233	seq_ID 188	<i>Paenibacillus</i> sp.	EDS49994.1	F	399
s991	seq_ID 189	<i>Paenibacillus</i> sp. IDR-2	ACS99948.1	F	399
s950	seq_ID 190	<i>Paenibacillus</i> sp. oral taxon 786	EES74793.1	F	428
s1280	seq_ID 191	<i>Paramecium tetraurelia</i>	145542269	F	400
s71	seq_ID 192	<i>Pelobacter carbinolicus</i>	ABA87701.1	F	494
s5	seq_ID 193	<i>Pelobacter carbinolicus</i>	ABA87615.1	F	435
s66	seq_ID 194	<i>Pelobacter propionicus</i>	ABK98395.1	F	486
s16	seq_ID 195	<i>Pelobacter propionicus</i>	ABK98811.1	F	467
s136	seq_ID 196	<i>Penicillium chrysogenum</i>	CAP99707.1	F	440
s936	seq_ID 197	<i>Planctomyces limnophilus</i>	EEO67214.1	F	490
s1158	seq_ID 198	<i>Planctomyces limnophilus</i>	EEO68341.1	F	412
s526	seq_ID 199	<i>Planctomyces maris</i>	EDL58855.1	F	392
s992	seq_ID 200	<i>Polypodiodes niponica</i>	BAI48071.1	Y	521
s942	seq_ID 201	<i>Polypodiodes niponica</i>	BAI48070.1	F	443
s1202	seq_ID 202	<i>Populus trichocarpa</i>	EEF12098.1	F	162
s168	seq_ID 203	<i>Ralstonia eutropha</i>	AAZ64302.1	F	452
s190	seq_ID 204	<i>Ralstonia eutropha</i>	CAJ96989.1	F	451
s81	seq_ID 205	<i>Ralstonia metallidurans</i>	ABF11015.1	F	448
s110	seq_ID 206	<i>Ralstonia metallidurans</i>	ABF11268.1	F	430
s123	seq_ID 207	<i>Rhizobium</i> sp.	P55348.1	F	433
s657	seq_ID 208	<i>Rhodopirellula baltica</i>	CAD74517.1	F	428
s4	seq_ID 209	<i>Rhodopseudomonas palustris</i>	ABJ08391.1	F	445
s130	seq_ID 210	<i>Rhodopseudomonas palustris</i>	CAA71101.1	F	433
s155	seq_ID 211	<i>Rhodopseudomonas palustris</i>	ABD06434.1	F	433
s97	seq_ID 212	<i>Rhodopseudomonas palustris</i>	ABD87279.1	F	433
s135	seq_ID 213	<i>Rhodopseudomonas palustris</i>	ACF02757.1	F	432
s84	seq_ID 214	<i>Rhodospirillum rubrum</i>	ABC20867.1	F	437
s1279	seq_ID 215	<i>Rubrobacter xylanophilus</i>	ABG05671.1	F	372
s915	seq_ID 216	<i>Saccharomonospora viridis</i>	ACU97316.1	F	428
s42	seq_ID 217	<i>Saccharopolyspora erythraea</i>	CAM03596.1	F	421
s82	seq_ID 218	<i>Schizosaccharomyces japonicus</i>	EEB08219.1	F	437
s923	seq_ID 219	<i>Sphaerobacter thermophilus</i>	ACZ39437.1	F	404
s924	seq_ID 220	<i>Streptomyces albus</i>	239983547	F	371
s23	seq_ID 221	<i>Streptomyces avermitilis</i>	BAC69361.1	F	450
s44	seq_ID 222	<i>Acaryochloris marina</i>	ABW29816.1	F	423
s921	seq_ID 223	<i>Streptomyces filamentosus</i>	239945642	F	447
s934	seq_ID 224	<i>Streptomyces flavogriseus</i>	EEW70811.1	F	447
s920	seq_ID 225	<i>Streptomyces ghanaensis</i>	239927462	F	448
s922	seq_ID 226	<i>Streptomyces griseoflavus</i>	256812310	F	448
s28	seq_ID 227	<i>Streptomyces griseus</i>	BAG17791.1	F	447
s926	seq_ID 228	<i>Streptomyces hygroscopicus</i>	256775136	F	414
s916	seq_ID 229	<i>Streptomyces lividans</i>	256783789	F	449
s33	seq_ID 230	<i>Streptomyces peucetius</i>	ACA52082.1	F	455
s27	seq_ID 231	<i>Streptomyces pristinaespiralis</i>	EDY61772.1	F	455
s933	seq_ID 232	<i>Streptomyces scabiei</i>	CBG68454.1	F	447
s37	seq_ID 233	<i>Streptomyces</i> sp.	EDX25760.1	F	453
s34	seq_ID 234	<i>Streptomyces</i> sp.	EDY46371.1	F	453
s931	seq_ID 235	<i>Streptomyces</i> sp. AA4	256668250	F	428
s918	seq_ID 236	<i>Streptomyces</i> sp. C	256770952	F	454
s929	seq_ID 237	<i>Streptomyces</i> sp. Mgl	254385931	F	453
s928	seq_ID 238	<i>Streptomyces</i> sp. SPB74	254379682	F	453
s930	seq_ID 239	<i>Streptomyces</i> sp. SPB78	256680470	F	404
s26	seq_ID 240	<i>Streptomyces viceus</i>	EDY55942.1	F	453
s927	seq_ID 241	<i>Streptomyces viridochromogenes</i>	256805984	F	447
s61	seq_ID 242	<i>Synechococcus</i> sp.	EDX84551.1	F	426

-continued

---

Program: CLUSTALW,  
 Default parameters:  
 Protein Gap Open Penalty 10.0  
 Protein Gap Extension Penalty 0.2  
 Protein weight matrix: Gonnet series

---

S_ID DB	SEQ ID NO	Organism	GI No. of the reference sequences	Aa	Position
s935	seq_ID 243	<i>Synechococcus</i> sp. PCC 7335	254422098	F	426
s53	seq_ID 244	<i>Synechocystis</i> sp.	BAA17978.1	F	428
s22	seq_ID 245	<i>Syntrophobacter fumaroxidans</i>	ABK18414.1	F	478
s6	seq_ID 246	<i>Syntrophobacter fumaroxidans</i>	ABK17672.1	F	457
s912	seq_ID 247	<i>Teredinibacter turnerae</i>	ACR13362.1	F	438
s57	seq_ID 248	<i>Thermosynechococcus elongatus</i>	BAC09861.1	F	425
s43	seq_ID 249	<i>Trichodesmium erythraeum</i>	ABG50159.1	F	418
s1178	seq_ID 250	Uncultured organism	ACA58560.1	F	118
s1176	seq_ID 251	Uncultured organism	ABL07557.1	F	118
s1165	seq_ID 252	Uncultured organism	ACA58559.1	F	116
s1166	seq_ID 253	Uncultured organism	ACA58558.1	F	116
s1168	seq_ID 254	Uncultured organism	ABL07560.1	F	116
s1169	seq_ID 255	Uncultured organism	ABL07565.1	F	116
s1170	seq_ID 256	Uncultured organism	ABL07566.1	F	116
s1167	seq_ID 257	Uncultured organism	ACA58545.1	F	116
s1171	seq_ID 258	Uncultured organism	ACA58535.1	F	116
s1180	seq_ID 259	Uncultured organism	ACA58549.1	F	116
s1179	seq_ID 260	Uncultured organism	ACA58554.1	F	116
s1181	seq_ID 261	Uncultured organism	ACA58555.1	F	116
s1182	seq_ID 262	Uncultured organism	ACA58556.1	F	116
s1235	seq_ID 263	Uncultured organism	ACA58530.1	F	116
s1188	seq_ID 264	Uncultured organism	ACA58534.1	F	115
s1237	seq_ID 265	Uncultured organism	ACA58552.1	F	115
s1223	seq_ID 266	Uncultured organism	ABL07558.1	F	115
s1200	seq_ID 267	Uncultured organism	ABL07542.1	F	115
s1236	seq_ID 268	Uncultured organism	ACA58539.1	F	114
s1238	seq_ID 269	Uncultured organism	ACA58537.1	F	114
s1233	seq_ID 270	Uncultured organism	ACA58543.1	F	114
s1173	seq_ID 271	Uncultured organism	ABL07553.1	F	114
s1241	seq_ID 272	Uncultured organism	ABL07540.1	F	114
s1242	seq_ID 273	Uncultured organism	ABL07544.1	F	114
s1225	seq_ID 274	Uncultured organism	ACA58557.1	F	114
s1183	seq_ID 275	Uncultured organism	ACA58520.1	F	113
s1197	seq_ID 276	Uncultured organism	ACA58524.1	F	113
s1185	seq_ID 277	Uncultured organism	ACA58522.1	F	113
s1190	seq_ID 278	Uncultured organism	ACA58525.1	F	113
s1187	seq_ID 279	Uncultured organism	ACA58523.1	F	113
s1184	seq_ID 280	Uncultured organism	ACA58521.1	F	113
s1204	seq_ID 281	Uncultured organism	ACA58547.1	F	113
s1221	seq_ID 282	Uncultured organism	ACA58544.1	F	113
s1198	seq_ID 283	Uncultured organism	ACA58546.1	F	112
s1226	seq_ID 284	Uncultured organism	ACA58527.1	F	112
s1227	seq_ID 285	Uncultured organism	ABL07537.1	F	112
s1232	seq_ID 286	Uncultured organism	ACA58510.1	F	112
s1230	seq_ID 287	Uncultured organism	ACA58538.1	F	112
s1229	seq_ID 288	Uncultured organism	ACA58542.1	F	112
s1231	seq_ID 289	Uncultured organism	ACA58540.1	F	112
s1207	seq_ID 290	Uncultured organism	ABL07564.1	F	112
s1212	seq_ID 291	Uncultured organism	ABL07563.1	F	112
s1208	seq_ID 292	Uncultured organism	ABL07562.1	F	112
s1209	seq_ID 293	Uncultured organism	ABL07559.1	F	112
s1214	seq_ID 294	Uncultured organism	ABL07556.1	F	112
s1216	seq_ID 295	Uncultured organism	ACA58528.1	F	112
s1219	seq_ID 296	Uncultured organism	ACA58536.1	F	112
s1192	seq_ID 297	Uncultured organism	ABL07533.1	F	112
s1195	seq_ID 298	Uncultured organism	ABL07536.1	F	112
s1174	seq_ID 299	Uncultured organism	ABL07545.1	F	112
s1186	seq_ID 300	Uncultured organism	ABL07548.1	F	112
s1196	seq_ID 301	Uncultured organism	ACA58561.1	F	112
s1172	seq_ID 302	Uncultured organism	ABL07555.1	F	112
s1194	seq_ID 303	Uncultured organism	ABL07541.1	F	112
s1211	seq_ID 304	Uncultured organism	ABL07554.1	F	112
s1220	seq_ID 305	Uncultured organism	ABL07547.1	F	112
s1203	seq_ID 306	Uncultured organism	ABL07550.1	F	112
s1199	seq_ID 307	Uncultured organism	ABL07551.1	F	112
s1228	seq_ID 308	Uncultured organism	ACA58509.1	F	111
s1201	seq_ID 309	Uncultured organism	ACA58514.1	F	111
s1205	seq_ID 310	Uncultured organism	ABL07543.1	F	111
s1206	seq_ID 311	Uncultured organism	ABL07534.1	F	111

---

Program: CLUSTALW,  
 Default parameters:  
 Protein Gap Open Penalty 10.0  
 Protein Gap Extension Penalty 0.2  
 Protein weight matrix: Gonnet series

---

S_ID DB	SEQ ID NO	Organism	GI No. of the reference sequences	Aa	Position
s1177	seq_ID 312	Uncultured organism	ABL07546.1	F	111
s1210	seq_ID 313	Uncultured organism	ABL07535.1	F	111
s1175	seq_ID 314	Uncultured organism	ABL07552.1	F	111
s1191	seq_ID 315	Uncultured organism	ABL07549.1	F	111
s1222	seq_ID 316	Uncultured organism	ACA58553.1	F	111
s1244	seq_ID 317	Uncultured organism	ABL07539.1	F	111
s1213	seq_ID 318	Uncultured organism	ACA58532.1	F	110
s1239	seq_ID 319	Uncultured organism	ACA58548.1	F	110
s1215	seq_ID 320	Uncultured organism	ABL07561.1	F	110
s1240	seq_ID 321	Uncultured organism	ACA58533.1	F	110
s1234	seq_ID 322	Uncultured organism	ABL07538.1	F	109
s1224	seq_ID 323	Uncultured organism	ACA58541.1	F	109
s1217	seq_ID 324	Uncultured organism	ACA58529.1	F	109
s596	seq_ID 325	<i>Verrucomicrobium spinosum</i>	171910093	F	395
s70	seq_ID 326	<i>Acidiphilium cryptum</i>	ABQ30890.1	F	430

---

Further potential cyclase mutants with the desired substrate properties can be produced starting from these, on the basis of the findings for mutants of Zm-SHC-1.

## 2. Further Proteins/Enzyme Mutants According to the Invention

The present invention is not limited to the mutants with cyclase activity concretely disclosed herein, but rather also extends to functional equivalents thereof.

“Functional equivalents” or analogs of the concretely disclosed enzymes and enzyme mutants (F486 and “F486-analog” mutants, derived from SEQ ID NO: 2 to 326, in particular SEQ ID NO: 2 to 6) are, within the scope of the present invention, various polypeptides thereof, which furthermore possess the desired biological activity, for example cyclase activity.

For example “functional equivalents” are understood to include enzymes and mutants that have, in a test applied for “cyclase activity” in the sense of the invention (i.e. with a reference substrate under standard conditions), an at least 1%, in particular at least about 5 to 10%, for example at least 10% or at least 20%, for example at least 50% or 75% or 90% higher or lower activity of an enzyme, comprising an amino acid sequence concretely defined herein (e.g. an F486 and “F486-analog” mutant, derived from SEQ ID NO: 2 to 326; in particular SEQ ID NO: 2 to 6).

The activity information for functional equivalents refers herein, unless stated otherwise, to activity determinations, performed by means of a reference substrate under standard conditions, as defined herein.

The “cyclase activity” in the sense of the invention can be detected by means of various known tests. Without being limited to this, we may mention a test using a reference substrate, for example citronellal racemate or R(+) form, under standard conditions, as described above and explained in the experimental section.

Functional equivalents are moreover stable e.g. between pH 4 to 11 and advantageously possess a pH optimum in a range from pH 5 to 10, such as in particular 6.5 to 9.5 or 7 to 8 or at about 7.5, and a temperature optimum in the range from 15° C. to 80° C. or 20° C. to 70° C., for example about 30 to 60° C. or about 35 to 45° C., such as at 40° C.

“Functional equivalents” are to be understood according to the invention to include in particular also “mutants”, which, as well as the concretely stated mutation(s) (e.g. an F486 and “F486-analog” mutant, derived from SEQ ID NO: 2 to 326, in particular SEQ ID NO: 2 to 6), have in at least one sequence position of the aforementioned amino acid sequences, an amino acid other than that concretely stated, but nevertheless possess one of the aforementioned biological activities.

“Functional equivalents” comprise the mutants obtainable by one or more, for example 1 to 50, 2 to 30, 2 to 15, 4 to 12 or 5 to 10 “additional mutations”, such as amino acid additions, substitutions, deletions and/or inversions, wherein the stated changes can occur in any sequence position, provided they lead to a mutant with the property profile according to the invention. Functional equivalence is in particular also present when the reactivity profiles between mutant and unaltered polypeptide coincide qualitatively, i.e. for example the same substrates are converted at a different rate.

“Additional mutations” of this kind occur at a position of the respective amino acid sequence different from position F486 according to SEQ ID NO: 2 or from the F486-analog position according to one of SEQ ID NOs: 3 to 326, in particular SEQ ID NO: 3 to 6.

Nonlimiting examples of suitable amino acid substitutions are given in the following table:

Original residue	Examples of substitution
Ala	Ser
Arg	Lys
Asn	Gln; His
Asp	Glu
Cys	Ser
Gln	Asn
Glu	Asp
Gly	Pro
His	Asn; Gln
Ile	Leu; Val
Leu	Ile; Val
Lys	Arg; Gln; Glu
Met	Leu; Ile

-continued

Original residue	Examples of substitution
Phe	Met; Leu; Tyr
Ser	Thr
Thr	Ser
Trp	Tyr
Tyr	Trp; Phe
Val	Ile; Leu

“Functional equivalents” in the above sense are also “precursors” of the polypeptides described as well as “functional derivatives” and “salts” of the polypeptides.

“Precursors” are natural or synthetic precursors of the polypeptides with or without the desired biological activity.

The term “salts” means both salts of carboxyl groups and salts of acid addition of amino groups of the protein molecules according to the invention. Salts of carboxyl groups can be produced in a manner known per se and comprise inorganic salts, for example sodium, calcium, ammonium, iron and zinc salts, and salts with organic bases, for example amines, such as triethanolamine, arginine, lysine, piperidine and the like. Salts of acid addition, for example salts with mineral acids, such as hydrochloric acid or sulfuric acid and salts with organic acids, such as acetic acid and oxalic acid, are also objects of the invention.

“Functional derivatives” of polypeptides according to the invention can also be produced on functional amino acid side groups or at their N- or C-terminal end by known techniques. Derivatives of this kind comprise for example aliphatic esters of carboxylic acid groups, amides of carboxylic acid groups, obtainable by reaction with ammonia or with a primary or secondary amine; N-acyl derivatives of free amino groups, produced by reaction with acyl groups; or O-acyl derivatives of free hydroxyl groups, produced by reaction with acyl groups.

“Functional equivalents” naturally also comprise polypeptides that are accessible from other organisms, and naturally occurring variants. For example areas of homologous sequence regions can be established by sequence comparison and equivalent enzymes can be determined based on the concrete information of the invention.

“Functional equivalents” also comprise fragments, preferably individual domains or sequence motifs, of the polypeptides according to the invention, which for example have the desired biological function.

“Functional equivalents” are moreover fusion proteins, which have one of the aforementioned polypeptide sequences or functional equivalents derived therefrom and at least one further, functionally different therefrom, heterologous sequence in functional N- or C-terminal linkage (i.e. without mutual substantial functional impairment of the fusion protein parts). Nonlimiting examples of heterologous sequences of this kind are e.g. signal peptides, histidine anchors or enzymes.

“Functional equivalents” that are also included according to the invention are homologs to the concretely disclosed proteins. These possess at least 60%, preferably at least 75%, especially at least 85%, for example 90, 91, 92, 93, 94, 95, 96, 97, 98 or 99%, homology (or identity) to one of the concretely disclosed amino acid sequences, calculated using the algorithm of Pearson and Lipman, Proc. Natl. Acad. Sci. (USA) 85(8), 1988, 2444-2448. A percentage homology or identity of a homologous polypeptide according to the invention means in particular percentage identity of the amino acid residues relative to the total length of one of the amino acid sequences concretely described herein. In par-

ticular, however, these homologs also have the F486 or “F486-analog” mutation, derived from SEQ ID NO:2 to 326, in particular SEQ ID NO: 2 to 6.

The percentage identity values can also be determined on the basis of BLAST alignments, blastp algorithms (protein-protein BLAST), or using the Clustal settings given below.

In the case of a possible protein glycosylation, “functional equivalents” according to the invention comprise proteins of the type designated above in deglycosylated or glycosylated form as well as modified forms obtainable by changing the glycosylation pattern.

Homologs of the proteins or polypeptides according to the invention can be produced by mutagenesis, e.g. by point mutation, lengthening or shortening of the protein.

Homologs of the proteins according to the invention can be identified by screening combinatorial databases of mutants, for example shortened mutants. For example a variegated database of protein variants can be produced by combinatorial mutagenesis at nucleic acid level, for example by enzymatic ligation of a mixture of synthetic oligonucleotides. There are a great many methods that can be used for producing databases of potential homologs from a degenerated oligonucleotide sequence. The chemical synthesis of a degenerated gene sequence can be carried out in an automatic DNA synthesizer, and the synthetic gene can then be ligated into a suitable expression vector. The use of a degenerated set of genes makes it possible to provide all sequences, in one mixture, which code for the desired set of potential protein sequences. Methods for the synthesis of degenerated oligonucleotides are known by a person skilled in the art (e.g. Narang, S. A. (1983) Tetrahedron 39:3; Itakura et al. (1984) Annu. Rev. Biochem. 53:323; Itakura et al., (1984) Science 198:1056; Ike et al. (1983) Nucleic Acids Res. 11:477).

Several techniques for screening gene products of combinatorial databases, which were produced by point mutations or shortening, and for screening cDNA databases for gene products with a chosen property, are known in the prior art. These techniques can be adapted for rapid screening of gene banks that have been produced by combinatorial mutagenesis of homologs according to the invention. The techniques used most often for screening large gene banks, as the basis for high-throughput analysis, comprise cloning the gene bank into replicatable expression vectors, transforming suitable cells with the resultant vector bank and expressing the combinatorial genes in conditions in which detection of the desired activity facilitates the isolation of the vector that codes for the gene whose product was detected. Recursive ensemble mutagenesis (REM), a technique that increases the frequency of functional mutants in the databases, can be used in combination with the screening tests, to identify homologs (Arkin and Yourvan (1992) PNAS 89:7811-7815; Delgrave et al. (1993) Protein Engineering 6(3):327-331).

### 3. Nucleic Acids and Constructs

#### 3.1 Nucleic Acids

The invention also relates to nucleic acid sequences that code for an enzyme as described above or a mutant thereof described above with cyclase activity.

The present invention also relates to nucleic acids with a specified degree of identity to the concrete sequences described herein.

“Identity” between two nucleic acids means identity of the nucleotides in each case over the whole length of nucleic acid, in particular the identity that is calculated by comparison by means of the Vector NTI Suite 7.1 software from the company Informax (USA) using the Clustal method (Higgins D G, Sharp P M. Fast and sensitive multiple sequence

alignments on a microcomputer. *Comput Appl. Biosci.* 1989 April; 5(2):151-1), setting the following parameters:  
Multiple Alignment Parameters:

Gap opening penalty	10
Gap extension penalty	10
Gap separation penalty range	8
Gap separation penalty	off
% identity for alignment delay	40
Residue specific gaps	off
Hydrophilic residue gap	off
Transition weighting	0

Pairwise Alignment Parameter:

FAST algorithm	on
K-tuple size	1
Gap penalty	3
Window size	5
Number of best diagonals	5

As an alternative, the identity can also be determined according to Chenna, Ramu, Sugawara, Hideaki, Koike, Tadashi, Lopez, Rodrigo, Gibson, Toby J, Higgins, Desmond G, Thompson, Julie D. Multiple sequence alignment with the Clustal series of programs. (2003) *Nucleic Acids Res* 31 (13):3497-500, according to Internet address: [ebi.ac.uk/Tools/clustalw/index.html#](http://ebi.ac.uk/Tools/clustalw/index.html#) and with the following parameters:

DNA Gap Open Penalty	15.0
DNA Gap Extension Penalty	6.66
DNA Matrix	Identity
Protein Gap Open Penalty	10.0
Protein Gap Extension Penalty	0.2
Protein matrix	Gonnet
Protein/DNA ENDGAP	-1
Protein/DNA GAPDIST	4

All nucleic acid sequences mentioned herein (single-stranded and double-stranded DNA and RNA sequences, for example cDNA and mRNA) can be produced in a manner known per se by chemical synthesis from the nucleotide building blocks, for example by fragment condensation of individual overlapping, complementary nucleic acid building blocks of the double helix. The chemical synthesis of oligonucleotides can for example be carried out in a known manner, by the phosphoroamidite technique (Voet, Voet, 2nd edition, Wiley Press New York, pages 896-897). The adding-on of synthetic oligonucleotides and filling of gaps using the Klenow fragment of DNA polymerase and ligation reactions as well as general cloning techniques are described in Sambrook et al. (1989), *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press.

The invention also relates to nucleic acid sequences (single-stranded and double-stranded DNA and RNA sequences, for example cDNA and mRNA), coding for one of the above polypeptides and functional equivalents thereof, which are accessible e.g. using artificial nucleotide analogs.

The invention relates both to isolated nucleic acid molecules, which code for polypeptides or proteins according to the invention or biologically active segments thereof, and to nucleic acid fragments, which can be used for example as hybridization probes or primers for the identification or amplification of coding nucleic acids according to the invention.

The nucleic acid molecules according to the invention can in addition contain untranslated sequences of the 3'- and/or 5'-end of the coding gene region.

The invention further comprises the nucleic acid molecules complementary to the concretely described nucleotide sequences, or a segment thereof.

The nucleotide sequences according to the invention make it possible to produce probes and primers that can be used for the identification and/or cloning of homologous sequences in other cell types and organisms. Said probes or primers usually comprise a nucleotide sequence region which hybridizes under "stringent" conditions (see below) to at least about 12, preferably at least about 25, for example about 40, 50 or 75 successive nucleotides of a sense strand of a nucleic acid sequence according to the invention or of a corresponding antisense strand.

An "isolated" nucleic acid molecule is separate from other nucleic acid molecules that are present in the natural source of the nucleic acid, and moreover can be essentially free of other cellular material or culture medium, when it is produced by recombinant techniques, or free of chemical precursors or other chemicals, when it is chemically synthesized.

A nucleic acid molecule according to the invention can be isolated by standard techniques of molecular biology and the sequence information provided according to the invention. For example, cDNA can be isolated from a suitable cDNA-bank, using one of the concretely disclosed complete sequences or a segment thereof as hybridization probe and standard hybridization techniques (as described for example in Sambrook, J., Fritsch, E. F. and Maniatis, T. *Molecular Cloning: A Laboratory Manual*. 2nd edition, Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989). Moreover, a nucleic acid molecule, comprising one of the disclosed sequences or a segment thereof, can be isolated by polymerase chain reaction, using the oligonucleotide primers that were constructed on the basis of this sequence. The nucleic acid thus amplified can be cloned into a suitable vector and can be characterized by DNA sequence analysis. The oligonucleotides according to the invention can moreover be produced by standard methods of synthesis, e.g. with an automatic DNA synthesizer.

Nucleic acid sequences according to the invention or derivatives thereof, homologs or parts of these sequences, can be isolated for example with usual hybridization methods or PCR techniques from other bacteria, e.g. via genomic or cDNA databases. These DNA sequences hybridize under standard conditions to the sequences according to the invention.

"Hybridization" means the capacity of a poly- or oligonucleotide to bind to an almost complementary sequence under standard conditions, whereas under these conditions nonspecific binding between noncomplementary partners does not occur. For this, the sequences can be up to 90-100% complementary. The property of complementary sequences of being able to bind specifically to one another is utilized for example in Northern or Southern blotting or in primer binding in PCR or RT-PCR.

Short oligonucleotides of the conserved regions are used advantageously for hybridization. However, longer fragments of the nucleic acids according to the invention or the complete sequences can also be used for hybridization. These standard conditions vary depending on the nucleic acid used (oligonucleotide, longer fragment or complete sequence) or depending on which type of nucleic acid, DNA or RNA, is used for hybridization. Thus, for example, the

melting temperatures for DNA:DNA hybrids are approx. 10° C. lower than those of DNA:RNA hybrids of the same length.

Standard conditions mean for example, depending on the nucleic acid, temperatures between 42 and 58° C. in an aqueous buffer solution with a concentration between 0.1 to 5×SSC (1×SSC=0.15 M NaCl, 15 mM sodium citrate, pH 7.2) or additionally in the presence of 50% formamide, for example 42° C. in 5×SSC, 50% formamide. Advantageously, the hybridization conditions for DNA:DNA hybrids are 0.1×SSC and temperatures between about 20° C. to 45° C., preferably between about 30° C. to 45° C. For DNA:RNA hybrids the hybridization conditions are advantageously 0.1×SSC and temperatures between about 30° C. to 55° C., preferably between about 45° C. to 55° C. These stated temperatures for hybridization are for example calculated melting temperature values for a nucleic acid with a length of approx. 100 nucleotides and a G+C content of 50% in the absence of formamide. The experimental conditions for DNA hybridization are described in relevant textbooks on genetics, for example Sambrook et al., "Molecular Cloning", Cold Spring Harbor Laboratory, 1989, and can be calculated using formulas known by a person skilled in the art, for example depending on the length of the nucleic acids, the type of hybrids or the G+C content. Further information on hybridization can be obtained by a person skilled in the art from the following textbooks: Ausubel et al. (eds), 1985, *Current Protocols in Molecular Biology*, John Wiley & Sons, New York; Hames and Higgins (eds), 1985, *Nucleic Acids Hybridization: A Practical Approach*, IRL Press at Oxford University Press, Oxford; Brown (ed), 1991, *Essential Molecular Biology: A Practical Approach*, IRL Press at Oxford University Press, Oxford.

"Hybridization" can in particular take place under stringent conditions. Said hybridization conditions are described for example by Sambrook, J., Fritsch, E. F., Maniatis, T. in: *Molecular Cloning (A Laboratory Manual)*, 2nd edition, Cold Spring Harbor Laboratory Press, 1989, pages 9.31-9.57 or in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6.

"Stringent" hybridization conditions mean in particular: Incubation at 42° C. overnight in a solution consisting of 50% formamide, 5×SSC (750 mM NaCl, 75 mM trisodium citrate), 50 mM sodium phosphate (pH7.6), 5×Denhardt solution, 10% dextran sulfate and 20 g/ml denatured, sheared salmon sperm DNA, followed by a step of washing the filters with 0.1×SSC at 65° C.

The invention also relates to derivatives of the concretely disclosed or derivable nucleic acid sequences.

Thus, further nucleic acid sequences according to the invention coding for cyclase mutants can be derived e.g. from SEQ ID NO: 1 or from the coding sequences for SEQ ID NO: 2 to 326, in particular SEQ ID NO: 2 to 6, by an F486 or F486-analog mutation and differ from them by addition, substitution, insertion or deletion of single or several nucleotides, but furthermore code for polypeptides with the desired property profile.

The invention also includes nucleic acid sequences that comprise so-called silent mutations or are altered corresponding to the codon-usage of a special original or host organism, compared with a concretely stated sequence, as well as naturally occurring variants, for example splice variants or allele variants, thereof.

It also relates to sequences obtainable by conservative nucleotide substitutions (i.e. the amino acid in question is replaced with an amino acid of the same charge, size, polarity and/or solubility).

The invention also relates to the molecules derived by sequence polymorphisms from the concretely disclosed nucleic acids. These genetic polymorphisms can exist between individuals within a population owing to natural variation. These natural variations usually bring about a variance of 1 to 5% in the nucleotide sequence of a gene.

Derivatives of the nucleic acid sequences according to the invention coding for cyclase mutants derived from sequence SEQ ID NO: 1 or from one of the coding sequences for SEQ ID NO: 2 to 326, in particular SEQ ID NO: 2 to 6, include for example allele variants that have at least 60% homology at the derived amino acid level, preferably at least 80% homology, quite especially preferably at least 90% homology over the whole sequence region (regarding homology at the amino acid level, reference should be made to the above account relating to polypeptides). The homologies can advantageously be higher over partial regions of the sequences.

Furthermore, derivatives also mean homologs of the nucleic acid sequences according to the invention, for example fungal or bacterial homologs, shortened sequences, single-strand DNA or RNA of the coding and noncoding DNA sequence.

Moreover, derivatives mean for example fusions with promoters. The promoters, which are added to the given nucleotide sequences, can be altered by at least one nucleotide exchange, at least one insertion, inversion and/or deletion, without the functionality or efficacy of the promoters being impaired. Moreover, the efficacy of the promoters can be increased by altering their sequence or they can be exchanged completely for more effective promoters even of organisms of a different species.

### 3.2 Generation of Functional Mutants

Furthermore, methods for producing functional mutants of enzymes according to the invention are known by a person skilled in the art.

Depending on the technology used, a person skilled in the art can introduce completely random or even more-directed mutations in genes or also noncoding nucleic acid regions (which for example are important for the regulation of expression) and then prepare gene libraries. The necessary methods of molecular biology are known by a person skilled in the art and for example are described in Sambrook and Russell, *Molecular Cloning*, 3rd edition, Cold Spring Harbor Laboratory Press 2001.

Methods for altering genes and therefore for altering the proteins that they encode have long been familiar to a person skilled in the art, for example

site-directed mutagenesis, in which single or several nucleotides of a gene are deliberately exchanged (Trower M K (Ed.) 1996; *In vitro mutagenesis protocols*. Humana Press, New Jersey),

saturation mutagenesis, in which a codon for any amino acid can be exchanged or added at any point of a gene (Kegler-Ebo D M, Docktor C M, DiMaio D (1994) *Nucleic Acids Res* 22:1593; Baretino D, Feigenbutz M, Valcárel R, Stunnenberg H G (1994) *Nucleic Acids Res* 22:541; Barik S (1995) *Mol Biotechnol* 3:1),

the error-prone polymerase chain reaction (error-prone PCR), in which nucleotide sequences are mutated by error-prone DNA polymerases (Eckert K A, Kunkel T A (1990) *Nucleic Acids Res* 18:3739);

the SeSaM method (sequence saturation method), in which preferred exchanges are prevented by the polymerase. Schenk et al., *Biospektrum*, Vol. 3, 2006, 277-279

the passaging of genes in mutator strains, in which, for example owing to defective DNA repair mechanisms, there is an increased mutation rate of nucleotide sequences (Greener A, Callahan M, Jerpseth B (1996) An efficient random mutagenesis technique using an *E. coli* mutator strain. In: Trower M K (Ed.) In vitro mutagenesis protocols. Humana Press, New Jersey), or DNA shuffling, in which a pool of closely related genes is formed and digested and the fragments are used as templates for a polymerase chain reaction, in which, by repeated strand separation and bringing together again, finally mosaic genes of full length are produced (Stemmer W P C (1994) Nature 370:389; Stemmer W P C (1994) Proc Natl Acad Sci USA 91:10747).

Using so-called directed evolution (described for instance in Reetz M T and Jaeger K-E (1999), Topics Curr Chem 200:31; Zhao H, Moore J C, Volkov A A, Arnold F H (1999), Methods for optimizing industrial enzymes by directed evolution, in: Demain A L, Davies J E (Ed.) Manual of industrial microbiology and biotechnology. American Society for Microbiology), a person skilled in the art can produce functional mutants in a directed manner and on a large scale. For this, in a first step, gene libraries of the respective proteins are first produced, for example using the methods given above. The gene libraries are expressed in a suitable way, for example by bacteria or by phage display systems.

The relevant genes of host organisms that express functional mutants with properties that largely correspond to the desired properties can be submitted to another round of mutation. The steps of mutation and selection or screening can be repeated iteratively until the present functional mutants have the desired properties to a sufficient extent. Using this iterative procedure, a limited number of mutations, for example 1, 2, 3, 4 or 5 mutations, can be effected in stages and can be assessed and selected for their influence on the enzyme property in question. The selected mutant can then be submitted to a further mutation step in the same way. In this way the number of individual mutants to be investigated can be reduced significantly.

The results according to the invention also provide important information relating to structure and sequence of the relevant enzymes, which is required for deliberately generating further enzymes with desired modified properties. In particular so-called "hot spots" can be defined, i.e. sequence segments that are potentially suitable for modifying an enzyme property by introducing targeted mutations.

Information can also be deduced regarding amino acid sequence positions, in the region of which mutations can be carried out that should probably have little effect on enzyme activity, and can be designated as potential "silent mutations".

### 3.3 Constructs

The invention further relates to, in particular recombinant, expression constructs, containing, under the genetic control of regulatory nucleic acid sequences, a nucleic acid sequence coding for a polypeptide according to the invention; and, in particular recombinant, vectors, comprising at least one of these expression constructs.

An "expression unit" means, according to the invention, a nucleic acid with expression activity, which comprises a promoter, as defined herein, and after functional linkage with a nucleic acid to be expressed or a gene, regulates the expression, i.e. the transcription and the translation of said nucleic acid or said gene. Therefore in this connection it is also called a "regulatory nucleic acid sequence". In addition to the promoter, other regulatory elements, for example enhancers, can also be present.

An "expression cassette" or "expression construct" means, according to the invention, an expression unit that is functionally linked to the nucleic acid to be expressed or the gene to be expressed. In contrast to an expression unit, an expression cassette therefore comprises not only nucleic acid sequences that regulate transcription and translation, but also the nucleic acid sequences that are to be expressed as protein as a result of the transcription and translation.

The terms "expression" or "overexpression" describe, in the context of the invention, the production or increase in intracellular activity of one or more enzymes in a microorganism, which are encoded by the corresponding DNA. For this, it is possible for example to introduce a gene into an organism, replace an existing gene with another gene, increase the copy number of the gene or genes, use a strong promoter or use a gene that codes for a corresponding enzyme with a high activity; optionally, these measures can be combined.

Preferably said constructs according to the invention comprise a promoter 5'-upstream of the respective coding sequence and a terminator sequence 3'-downstream and optionally other usual regulatory elements, in each case operatively linked with the coding sequence.

A "promoter, of a "nucleic acid with promoter activity" or of a "promoter sequence" means, according to the invention, a nucleic acid which, functionally linked to a nucleic acid to be transcribed, regulates the transcription of said nucleic acid.

A "functional" or "operative" linkage means, in this connection, for example the sequential arrangement of one of the nucleic acids with promoter activity and of a nucleic acid sequence to be transcribed and optionally further regulatory elements, for example nucleic acid sequences that ensure the transcription of nucleic acids, and for example a terminator, in such a way that each of the regulatory elements can perform its function during transcription of the nucleic acid sequence. This does not necessarily require a direct linkage in the chemical sense. Genetic control sequences, for example enhancer sequences, can even exert their function on the target sequence from more remote positions or even from other DNA molecules. Arrangements are preferred in which the nucleic acid sequence to be transcribed is positioned behind (i.e. at the 3'-end of) the promoter sequence, so that the two sequences are joined together covalently. The distance between the promoter sequence and the nucleic acid sequence to be expressed transgenically can be smaller than 200 base pairs, or smaller than 100 base pairs or smaller than 50 base pairs.

In addition to promoters and terminator, the following may be mentioned as examples of other regulatory elements: targeting sequences, enhancers, polyadenylation signals, selectable markers, amplification signals, replication origins and the like. Suitable regulatory sequences are described for example in Goeddel, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, Calif. (1990).

Nucleic acid constructs according to the invention comprise in particular a sequence coding for a cyclase mutant, e.g. derived from SEQ ID NO: 1 or coding for a mutant of SEQ ID NO: 2 to 326 or derivatives and homologs thereof, and the nucleic acid sequences derivable therefrom, which have been linked operatively or functionally with one or more regulatory signals advantageously for controlling, e.g. increasing, gene expression.

In addition to these regulatory sequences, the natural regulation of these sequences can still be present before the actual structural genes and optionally can have been geneti-

cally altered, so that the natural regulation has been switched off and expression of the genes has been increased. The nucleic acid construct can, however, also be of simpler construction, i.e. no additional regulatory signals have been inserted before the coding sequence and the natural promoter, with its regulation, has not been removed. Instead, the natural regulatory sequence is mutated so that regulation no longer takes place and gene expression is increased.

A preferred nucleic acid construct advantageously also contains one or more of the "enhancer" sequences already mentioned, functionally linked to the promoter, which make increased expression of the nucleic acid sequence possible. Additional advantageous sequences can also be inserted at the 3'-end of the DNA sequences, such as further regulatory elements or terminators. One or more copies of the nucleic acids according to the invention can be contained in the construct. The construct can also contain other markers, such as antibiotic resistances or auxotrophy complementing genes, optionally for selection on the construct.

Examples of suitable regulatory sequences are contained in promoters such as *cos*-, *tac*-, *trp*-, *tet*-, *trp-tet*-, *lpp*-, *lac*-, *lpp-lac*-, *lacI<sup>q</sup>*-, *T7*-, *T5*-, *T3*-, *gal*-, *trc*-, *ara*-, *rhaP* (*rhaP<sub>BAD</sub>*) *SP6*-, *lambda-P<sub>R</sub>*- or in the *lambda-P<sub>L</sub>*-promoter, which advantageously find application in gram-negative bacteria. Further advantageous regulatory sequences are contained for example in the gram-positive promoters *amy* and *SPO2*, in the yeast or fungal promoters *ADC1*, *MFalpha*, *AC*, *P-60*, *CYC1*, *GAPDH*, *TEF*, *rp28*, *ADH*. Artificial promoters can also be used for regulation.

For expression in a host organism, the nucleic acid construct is advantageously inserted into a vector, for example a plasmid or a phage, which makes optimal expression of the genes in the host possible. Apart from plasmids and phage, vectors are also to be understood as all other vectors known by a person skilled in the art, e.g. viruses, such as SV40, CMV, baculovirus and adenovirus, transposons, IS elements, phasmids, cosmids, and linear or circular DNA. These vectors can be replicated autonomously in the host organism or can be replicated chromosomally. These vectors represent a further embodiment of the invention.

Suitable plasmids are for example in *E. coli* pLG338, pACYC184, pBR322, pUC18, pUC19, pKC30, pRep4, pHS1, pKK223-3, pDHE19.2, pHS2, pPlc236, pMBL24, pLG200, pUR290, pIN-III<sup>113</sup>-B1,  $\lambda$ gt11 or pBdCl, in *Streptomyces* pIJ101, pIJ364, pIJ702 or pIJ361, in *Bacillus* pUB110, pC194 or pBD214, in *Corynebacterium* pSA77 or pAJ667, in fungi pALS1, pIL2 or pBB116, in yeasts 2alphaM, pAG-1, YEp6, YEp13 or pEMBLye23 or in plants pLGV23, pGHlac<sup>+</sup>, pBIN19, pAK2004 or pDH51. The stated plasmids represent a small selection of the possible plasmids. Further plasmids are well known by a person skilled in the art and can for example be found in the book Cloning Vectors (Eds. Pouwels P. H. et al. Elsevier, Amsterdam-New York-Oxford, 1985, ISBN 0 444 904018).

In another embodiment of the vector, the vector containing the nucleic acid construct according to the invention or the nucleic acid according to the invention can also advantageously be introduced in the form of a linear DNA into the microorganisms and integrated via heterologous or homologous recombination into the genome of the host organism. This linear DNA can consist of a linearized vector such as a plasmid or only of the nucleic acid construct or the nucleic acid according to the invention.

For optimal expression of heterologous genes in organisms, it is advantageous to alter the nucleic acid sequences corresponding to the specific "codon usage" used in the

organism. The "codon usage" can easily be determined on the basis of computer evaluations of other known genes of the organism in question.

An expression cassette according to the invention is produced by fusion of a suitable promoter with a suitable coding nucleotide sequence and a terminator signal or polyadenylation signal. Common recombination and cloning techniques are used, as described for example in T. Maniatis, E. F. Fritsch and J. Sambrook, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. (1989) and in T. J. Silhavy, M. L. Berman and L. W. Enquist, Experiments with Gene Fusions, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. (1984) and in Ausubel, F. M. et al., Current Protocols in Molecular Biology, Greene Publishing Assoc. and Wiley Interscience (1987).

For expression in a suitable host organism, advantageously the recombinant nucleic acid construct or gene construct is inserted into a host-specific vector, which makes optimal expression of the genes in the host possible. Vectors are well known by a person skilled in the art and are given for example in "Cloning vectors" (Pouwels P. H. et al., Ed., Elsevier, Amsterdam-New York-Oxford, 1985).

#### 4. Microorganisms

Depending on the context, the term "microorganism" can mean the wild-type microorganism or a genetically altered, recombinant microorganism or both.

Using the vectors according to the invention, recombinant microorganisms can be produced, which are for example transformed with at least one vector according to the invention and can be used for producing the polypeptides according to the invention. Advantageously, the recombinant constructs according to the invention, described above, are introduced into a suitable host system and expressed. Preferably common cloning and transfection methods, known by a person skilled in the art, are used, for example coprecipitation, protoplast fusion, electroporation, retroviral transfection and the like, for expressing the stated nucleic acids in the respective expression system. Suitable systems are described for example in Current Protocols in Molecular Biology, F. Ausubel et al., Ed., Wiley Interscience, New York 1997, or Sambrook et al. Molecular Cloning: A Laboratory Manual. 2nd edition, Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.

In principle, all prokaryotic or eukaryotic organisms may be considered as recombinant host organisms for the nucleic acid according to the invention or the nucleic acid construct. Advantageously, microorganisms such as bacteria, fungi or yeasts are used as host organisms.

Advantageously, gram-positive or gram-negative bacteria are used, preferably bacteria of the families Enterobacteriaceae, Pseudomonadaceae, Rhizobiaceae, Streptomycetaceae or Nocardiaceae, especially preferably bacteria of the genera *Escherichia*, *Pseudomonas*, *Streptomyces*, *Nocardia*, *Burkholderia*, *Salmonella*, *Agrobacterium*, *Clostridium* or *Rhodococcus*. The genus and species *Escherichia coli* is quite especially preferred. Furthermore, other advantageous bacteria are to be found in the group of alpha-Proteobacteria, beta-Proteobacteria or gamma-Proteobacteria.

The host organism or the host organisms according to the invention preferably contain at least one of the nucleic acid sequences, nucleic acid constructs or vectors described in the present invention, which code for an enzyme with phenylethanol dehydrogenase activity according to the above definition.

Depending on the host organism, the organisms used in the method according to the invention are grown or cultured in a manner known by a person skilled in the art. Microorganisms are as a rule grown in a liquid medium, which contains a carbon source generally in the form of sugars, a nitrogen source generally in the form of organic nitrogen sources such as yeast extract or salts such as ammonium sulfate, trace elements such as iron, manganese and magnesium salts and optionally vitamins, at temperatures between 0° C. and 100° C., preferably between 10° C. to 60° C. with oxygen aeration. The pH of the liquid nutrient can be kept at a fixed value, i.e. regulated or not during culture. Culture can be batchwise, semi-batchwise or continuous. Nutrients can be present at the beginning of fermentation or can be supplied later, semicontinuously or continuously.

### 5. Recombinant Production of Enzymes According to the Invention

The invention further relates to methods for recombinant production of polypeptides according to the invention or functional, biologically active fragments thereof, wherein a polypeptide-producing microorganism is cultured, optionally the expression of the polypeptides is induced and these are isolated from the culture. The polypeptides can also be produced in this way on an industrial scale, if desired.

The microorganisms produced according to the invention can be cultured continuously or discontinuously in the batch method or in the fed-batch method or repeated fed-batch method. A summary of known cultivation methods can be found in the textbook by Chmiel (Bioprozesstechnik 1. Einführung in die Bioverfahrenstechnik [Bioprocess technology 1. Introduction to bioprocess technology] (Gustav Fischer Verlag, Stuttgart, 1991)) or in the textbook by Storhas (Bioreaktoren und periphere Einrichtungen [Bioreactors and peripheral equipment] (Vieweg Verlag, Braunschweig/Wiesbaden, 1994)).

The culture medium to be used must suitably meet the requirements of the respective strains. Descriptions of culture media for various microorganisms are given in the manual "Manual of Methods for General Bacteriology" of the American Society for Bacteriology (Washington D. C., USA, 1981).

These media usable according to the invention usually comprise one or more carbon sources, nitrogen sources, inorganic salts, vitamins and/or trace elements.

Preferred carbon sources are sugars, such as mono-, di- or polysaccharides. Very good carbon sources are for example glucose, fructose, mannose, galactose, ribose, sorbose, ribulose, lactose, maltose, sucrose, raffinose, starch or cellulose. Sugars can also be added to the media via complex compounds, such as molasses, or other by-products of sugar refining. It can also be advantageous to add mixtures of different carbon sources. Other possible carbon sources are oils and fats, for example soybean oil, sunflower oil, peanut oil and coconut oil, fatty acids, for example palmitic acid, stearic acid or linoleic acid, alcohols, for example glycerol, methanol or ethanol and organic acids, for example acetic acid or lactic acid.

Nitrogen sources are usually organic or inorganic nitrogen compounds or materials that contain these compounds. Examples of nitrogen sources comprise ammonia gas or ammonium salts, such as ammonium sulfate, ammonium chloride, ammonium phosphate, ammonium carbonate or ammonium nitrate, nitrates, urea, amino acids or complex nitrogen sources, such as corn-steep liquor, soya flour, soya protein, yeast extract, meat extract and others. The nitrogen sources can be used alone or as a mixture.

Inorganic salt compounds that can be present in the media comprise the chloride, phosphorus or sulfate salts of calcium, magnesium, sodium, cobalt, molybdenum, potassium, manganese, zinc, copper and iron.

Inorganic sulfur-containing compounds, for example sulfates, sulfites, dithionites, tetrathionates, thiosulfates, sulfides, as well as organic sulfur compounds, such as mercaptans and thiols, can be used as the sulfur source.

Phosphoric acid, potassium dihydrogen phosphate or dipotassium hydrogen phosphate or the corresponding sodium-containing salts can be used as the phosphorus source.

Chelating agents can be added to the medium, in order to keep the metal ions in solution. Especially suitable chelating agents comprise dihydroxyphenols, such as catechol or protocatechuic acid, or organic acids, such as citric acid.

The fermentation media used according to the invention usually also contain other growth factors, such as vitamins or growth promoters, which include for example biotin, riboflavin, thiamine, folic acid, nicotinic acid, pantothenate and pyridoxine. Growth factors and salts often originate from the components of complex media, such as yeast extract, molasses, corn-steep liquor and the like. Moreover, suitable precursors can be added to the culture medium. The exact composition of the compounds in the medium is strongly dependent on the respective experiment and is decided for each specific case individually. Information on media optimization can be found in the textbook "Applied Microbiol. Physiology, A Practical Approach" (Ed. P. M. Rhodes, P. F. Stanbury, IRL Press (1997) p. 53-73, ISBN 0 19 963577 3). Growth media can also be obtained from commercial suppliers, such as Standard 1 (Merck) or BHI (brain heart infusion, DIEGO) and the like.

All components of the medium are sterilized, either by heat (20 min at 1.5 bar and 121° C.) or by sterile filtration. The components can either be sterilized together, or separately if necessary. All components of the medium can be present at the start of culture or can be added either continuously or batchwise.

The culture temperature is normally between 15° C. and 45° C., preferably 25° C. to 40° C. and can be varied or kept constant during the experiment. The pH of the medium should be in the range from 5 to 8.5, preferably around 7.0. The pH for growing can be controlled during growing by adding basic compounds such as sodium hydroxide, potassium hydroxide, ammonia or ammonia water or acid compounds such as phosphoric acid or sulfuric acid. Antifoaming agents, for example fatty acid polyglycol esters, can be used for controlling foaming. To maintain the stability of plasmids, suitable selective substances, for example antibiotics, can be added to the medium. To maintain aerobic conditions, oxygen or oxygen-containing gas mixtures, for example ambient air, are fed into the culture. The temperature of the culture is normally in the range from 20° C. to 45° C. The culture is continued until a maximum of the desired product has formed. This target is normally reached within 10 hours to 160 hours.

The fermentation broth is then processed further. Depending on requirements, the biomass can be removed from the fermentation broth completely or partially by separation techniques, for example centrifugation, filtration, decanting or a combination of these methods or can be left in it completely.

If the polypeptides are not secreted in the culture medium, the cells can also be lysed and the product can be obtained from the lysate by known methods for isolation of proteins. The cells can optionally be disrupted with high-frequency

ultrasound, high pressure, for example in a French press, by osmolytic, by the action of detergents, lytic enzymes or organic solvents, by means of homogenizers or by a combination of several of the aforementioned methods.

The polypeptides can be purified by known chromatographic techniques, such as molecular sieve chromatography (gel filtration), such as Q-sepharose chromatography, ion exchange chromatography and hydrophobic chromatography, and with other usual techniques such as ultrafiltration, crystallization, salting-out, dialysis and native gel electrophoresis. Suitable methods are described for example in Cooper, T. G., *Biochemische Arbeitsmethoden* [Biochemical processes], Verlag Walter de Gruyter, Berlin, New York or in Scopes, R., *Protein Purification*, Springer Verlag, New York, Heidelberg, Berlin.

For isolating the recombinant protein, it can be advantageous to use vector systems or oligonucleotides, which lengthen the cDNA by defined nucleotide sequences and therefore code for altered polypeptides or fusion proteins, which for example serve for easier purification. Suitable modifications of this type are for example so-called "tags" functioning as anchors, for example the modification known as hexa-histidine anchor or epitopes that can be recognized as antigens of antibodies (described for example in Harlow, E. and Lane, D., 1988, *Antibodies: A Laboratory Manual*. Cold Spring Harbor (N.Y.) Press). These anchors can serve for attaching the proteins to a solid carrier, for example a polymer matrix, which can for example be used as packing in a chromatography column, or can be used on a microtiter plate or on some other carrier.

At the same time these anchors can also be used for recognition of the proteins. For recognition of the proteins, it is moreover also possible to use usual markers, such as fluorescent dyes, enzyme markers, which form a detectable reaction product after reaction with a substrate, or radioactive markers, alone or in combination with the anchors for derivatization of the proteins.

For the expression of mutants according to the invention, reference may be made to the description of expression of the wild-type enzyme EbN1 and the expression systems usable for this in WO2005/108590 and WO2006/094945, to which reference is hereby expressly made.

#### 6. Enzyme Immobilization

The enzymes according to the invention can be used free or immobilized in the method described herein. An immobilized enzyme is an enzyme that is fixed to an inert carrier. Suitable carrier materials and the enzymes immobilized thereon are known from EP-A-1149849, EP-A-1 069 183 and DE-OS 100193773 and from the references cited therein. Reference is made in this respect to the disclosure of these documents in their entirety. Suitable carrier materials include for example clays, clay minerals, such as kaolinite, diatomaceous earth, perlite, silica, aluminum oxide, sodium carbonate, calcium carbonate, cellulose powder, anion exchanger materials, synthetic polymers, such as polystyrene, acrylic resins, phenol formaldehyde resins, polyurethanes and polyolefins, such as polyethylene and polypropylene. For making the supported enzymes, the carrier materials are usually employed in a finely-divided, particulate form, porous forms being preferred. The particle size of the carrier material is usually not more than 5 mm, in particular not more than 2 mm (particle-size distribution curve). Similarly, when using dehydrogenase as whole-cell catalyst, a free or immobilized form can be selected. Carrier materials are e.g. Ca-alginate, and carrageenan. Enzymes as well as cells can also be crosslinked directly with glutaraldehyde (cross-linking to CLEAs). Corresponding and other

immobilization techniques are described for example in J. Lalonde and A. Margolin "Immobilization of Enzymes" in K. Drauz and H. Waldmann, *Enzyme Catalysis in Organic Synthesis 2002*, Vol. III, 991-1032, Wiley-VCH, Weinheim. Further information on biotransformations and bioreactors for carrying out methods according to the invention are also given for example in Rehm et al. (Ed.) *Biotechnology*, 2nd Edn, Vol 3, Chapter 17, VCH, Weinheim.

#### 7. Enzymatic Cyclization of Terpenes

##### 7.1 General Description

In particular, the method of cyclization according to the invention is carried out in the presence of an enzyme, wherein the enzyme is encoded by a nucleic acid sequence according to SEQ ID NO: 1 or a functional equivalent thereof, wherein the nucleic acid sequence is a constituent of a gene construct or vector. Said gene constructs or vectors are described in detail in international application PCT/EP2010/057696 on pages 16 to 20, to which reference is expressly made here. Said functional equivalents, in particular those with citronellal-isopulegol cyclase activity, comprise in particular an F486 or F486-analog mutation, as defined herein.

The host cell, which contains a gene construct or a vector, in which the nucleic acid sequence is contained that codes for the enzyme with the desired activity, is also designated as transgenic organism. The production of said transgenic organisms is known in principle and is discussed for example in international application PCT/EP2010/057696 on page 20, to which reference is expressly made here.

Cells from the group comprising bacteria, cyanobacteria, fungi and yeasts are preferably selected as transgenic organisms. The cell is preferably selected from fungi of the genus *Pichia* or bacteria of the genera *Escherichia*, *Corynebacterium*, *Ralstonia*, *Clostridium*, *Pseudomonas*, *Bacillus*, *Zymomonas*, *Rhodobacter*, *Streptomyces*, *Burkholderia*, *Lactobacillus* or *Lactococcus*. Especially preferably, the cell is selected from bacteria of the species *Escherichia coli*, *Pseudomonas putida*, *Burkholderia glumae*, *Streptomyces lividans*, *Streptomyces coelicolor* or *Zymomonas mobilis*.

A method according to the invention is preferred, characterized in that the enzyme with the activity of a citronellal-isopulegol cyclase is encoded by a gene that was isolated from a microorganism, selected from *Zymomonas mobilis*, *Methylococcus capsulatus*, *Rhodospseudomonas palustris*, *Bradyrhizobium japonicum*, *Frankia spec*, *Streptomyces coelicolor* and *Acetobacter pasteurianus*. The relevant genes isolated from *Zymomonas mobilis*, *Streptomyces coelicolor*, *Bradyrhizobium japonicum* and *Acetobacter pasteurianus* should be mentioned in particular.

A method according to the invention is further preferred, characterized in that the enzyme with cyclase activity was generated by a microorganism that overproduces the enzyme and that was selected from the group of microorganisms comprising the genera *Escherichia*, *Corynebacterium*, *Ralstonia*, *Clostridium*, *Pseudomonas*, *Bacillus*, *Zymomonas*, *Rhodobacter*, *Streptomyces*, *Burkholderia*, *Lactobacillus* and *Lactococcus*.

In particular, a method according to the invention should be mentioned that is characterized in that the enzyme with cyclase activity was produced by transgenic microorganisms of the species *Escherichia coli*, *Pseudomonas putida*, *Burkholderia glumae*, *Corynebacterium glutamicum*, *Saccharomyces cerevisiae*, *Pichia pastoris*, *Streptomyces lividans*, *Streptomyces coelicolor*, *Bacillus subtilis* or *Zymomonas mobilis*, which overproduce the enzyme with cyclase activity.

Further embodiments for carrying out the biocatalytic cyclization method according to the invention, such as, for example, the method for production of isopulegol:

The method according to the invention is characterized in that the enzyme is in at least one of the following forms:

- a) free, optionally purified or partially purified polypeptide;
- b) immobilized polypeptide;
- c) polypeptide isolated from cells according to a) or b);
- d) whole cell, optionally dormant or growing cells, comprising at least one such polypeptide;
- e) lysate or homogenizate of the cells according to d).

Another embodiment of the method according to the invention is characterized in that the cells are microorganisms, preferably transgenic microorganisms expressing at least one heterologous nucleic acid molecule coding for a polypeptide with the cyclase activity.

A preferred embodiment of the method according to the invention comprises at least the following steps a), b) and d):

- a) isolating or recombinantly producing a microorganism producing an enzyme with cyclase activity from a natural source or,
- b) multiplying this microorganism,
- c) optionally isolating the enzyme with cyclase activity from the microorganism or preparing a protein fraction comprising said enzyme, and
- d) transferring the microorganism according to stage b) or the enzyme according to stage c) to a medium that contains substrate, e.g. citronellal of general formula (I).

In the method according to the invention, substrate, such as, for example, citronellal is contacted with the enzyme, that has the activity of a citronellal-isopulegol cyclase, in a medium and/or is incubated so that conversion of the substrate, such as, for example, of citronellal, to isopulegol, takes place in the presence of the enzyme. Preferably the medium is an aqueous reaction medium.

The pH of the aqueous reaction medium in which the method according to the invention is preferably carried out is advantageously maintained between pH 4 and 12, preferably between pH 4.5 and 9, especially preferably between pH 5 and 8.

The aqueous reaction media are preferably buffered solutions, which as a rule have a pH of preferably from 5 to 8. The buffer used can be a citrate, phosphate, TRIS (Tris (hydroxymethyl)-aminomethane) or MES buffer (2-(N-morpholino)ethanesulfonic acid). Moreover, the reaction medium can contain other additives, for example detergents (for example taurodeoxycholate).

The substrate, such as, for example, citronellal, is used preferably in a concentration of 2-200 mM, especially preferably 5-25 mM in the enzymatic reaction and can be supplied continuously or discontinuously.

As a rule the enzymatic cyclization takes place at a reaction temperature below the deactivation temperature of the enzyme used and above  $-10^{\circ}\text{C}$ . Preferably the method according to the invention is carried out at a temperature between  $0^{\circ}\text{C}$ . and  $95^{\circ}\text{C}$ ., especially preferably at a temperature between  $15^{\circ}\text{C}$ . and  $60^{\circ}\text{C}$ ., in particular between  $20^{\circ}\text{C}$ . and  $40^{\circ}\text{C}$ ., e.g. at about  $25^{\circ}\text{C}$  to  $30^{\circ}\text{C}$ .

A method according to the invention in which the reaction of citronellal to isopulegol takes place at a temperature in the range from  $20^{\circ}\text{C}$  to  $40^{\circ}\text{C}$ . and/or a pH in the range from 4 to 8 is especially preferred.

As well as these single-phase aqueous systems, in another variant of the invention, two-phase systems are also used. Then, as well as an aqueous phase, organic, non-water-miscible reaction media are used as the second phase. As a

result, the reaction products accumulate in the organic phase. After the reaction, the product, such as, for example, isopulegol, in the organic phase can easily be separated from the aqueous phase that comprises the biocatalyst.

A method according to the invention is preferred wherein the production of isopulegol takes place in single-phase aqueous systems or in two-phase systems.

The reaction product isopulegol can be extracted with organic solvents and optionally can be distilled for purification.

Suitable organic solvents are for example aliphatic hydrocarbons, preferably with 5 to 8 carbon atoms, such as pentane, cyclopentane, hexane, cyclohexane, heptane, octane or cyclooctane, halogenated aliphatic hydrocarbons, preferably with one or two carbon atoms, such as dichloromethane, chloroform, carbon tetrachloride, dichloroethane or tetrachloroethane, aromatic hydrocarbons, such as benzene, toluene, the xylenes, chlorobenzene or dichlorobenzene, aliphatic acyclic and cyclic ethers or alcohols, preferably with 4 to 8 carbon atoms, such as ethanol, isopropanol, diethyl ether, methyl-tert-butyl ether, ethyl-tert-butyl ether, dipropyl ether, diisopropyl ether, dibutyl ether, tetrahydrofuran or esters such as ethyl acetate or n-butyl acetate or ketones such as methyl isobutyl ketone or dioxane or mixtures thereof. Especially preferably, the aforementioned heptane, methyl-tert-butyl ether, diisopropyl ether, tetrahydrofuran, and ethyl acetate are used.

The cyclases used according to the invention can be used in the method according to the invention as free or immobilized enzyme, as already described above.

For the method according to the invention it is possible to use dormant or growing, free or immobilized cells, which contain nucleic acids, nucleic acid constructs or vectors coding for the cyclase. Lysed cells, such as cell lysates or cell homogenates can also be used. Lysed cells are for example cells that have been permeabilized by a treatment for example with solvents, or cells that have been disrupted by an enzyme treatment, by a mechanical treatment (e.g. French press or ultrasound) or by some other method. The resultant raw extracts are advantageously suitable for the method according to the invention. Purified or partially purified enzymes can also be used for the method.

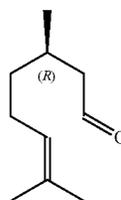
Where free organisms or enzymes are used for the method according to the invention, they are usefully isolated, via a filtration or centrifugation, for example, prior to the extraction.

The method according to the invention can be operated batchwise, semibatchwise or continuously.

## 7.2. Enzymatic Cyclization of Citronellal

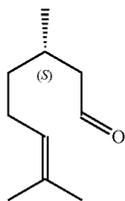
The citronellal of formula (II) used in accordance with the invention, and converted by means of an enzyme having citronellal-isopulegol cyclase activity, is available commercially both as (+)-R-citronellal of the formula (R-II) and as (-)-S-citronellal of the formula (S-II), and as a racemate of the formula (II).

(R-II)

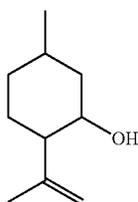


43

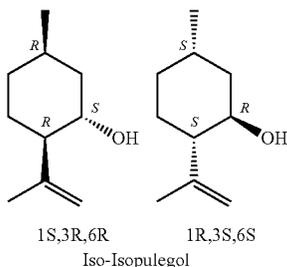
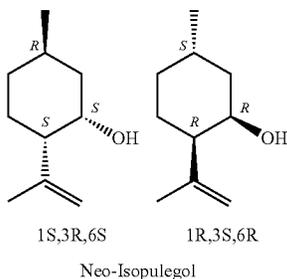
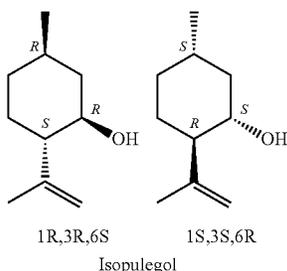
-continued



The isopulegol formed in accordance with the invention, of formula (I)

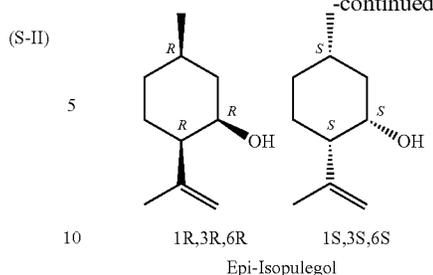


has a stereocenter in each of positions 1, 3 and 6, and so in principle there are 4 different diastereomers each with 2 enantiomers conceivable, in other words a total of 8 stereoisomers, if the starting point is the racemate of the citronellal of formula (I).



44

-continued



Suitable enzymes having the activity of a citronellal-isopulegol cyclase are intramolecular transferases from the subclass of the isomerases; that is, proteins having the enzyme code EC 5.4 (enzyme code in accordance with Eur. J. Biochem. 1999, 264, 610-650). Preferably they are representatives having the enzyme code 5.4.99.17. Also suitable in particular as enzymes having the activity of citronellal-isopulegol cyclase are those cyclases which also bring about the cyclization of homofarnesol to ambroxan or of squalene to hopene, which are described exhaustively in international application PCT/EP2010/057696, hereby incorporated by reference; the enzymes and mutants described here are also suitable.

One particularly suitable embodiment of the method according to the invention is that wherein the enzyme used in the method according to the invention and having the activity of a citronellal-isopulegol cyclase possesses a polypeptide sequence which either

- is SEQ ID NO: 2, or
- in which up to 25% of the amino acid residues are altered relative to SEQ ID NO: 2 by deletion, insertion, substitution or a combination thereof, and which still has at least 50% of the enzymatic activity of SEQ ID NO: 2.

Suitable enzymes with citronellal-isopulegol cyclase activity and comprising an amino sequence according to SEQ ID NO: 2, and also "functional equivalents" or analogs of the specifically disclosed enzymes (E) having citronellal-isopulegol cyclase activity, are described, as already indicated above, exhaustively in the international application PCT/EP2010/057696, hereby incorporated by reference.

In one particularly preferred embodiment of the method, the enzyme having citronellal-isopulegol cyclase activity is selected from enzymes which comprise an amino acid sequence according to SEQ ID NO: 2 or a sequence derived therefrom in which up to 25%, preferably up to 20%, more preferably up to 15%, in particular up to 10, 9, 8, 7, 6, 5, 4, 3, 2, 1% of the amino acid residues have been altered by a deletion, a substitution, an insertion or a combination of deletion, substitution and insertion, the polypeptide sequences altered relative to SEQ ID NO: 2 still possessing at least 50%, preferably 65%, more preferably 80%, more particularly more than 90% of the enzymatic activity of SEQ ID NO: 2. In this context, enzymatic activity of SEQ ID NO: 2 refers to the capacity to effect biocatalytic cyclization of citronellal of general formula (II) to the corresponding isopulegol of formula (I).

The method according to the invention is carried out preferably in the presence of an enzyme, the enzyme being encoded by a nucleic acid sequence according to SEQ ID NO: 1 or a functional equivalent thereof.

Functional equivalents here describe in principle nucleic acid sequences which under standard conditions undergo hybridization with a nucleic acid sequence or parts of a nucleic acid sequence and are capable of bringing about the

expression of a protein having the same properties as those of the enzyme having citronellal-isopulegol cyclase activity in a cell or in an organism.

A functional equivalent is additionally understood to refer to nucleic acid sequences which are homologous or identical to a defined percentage with a particular nucleic acid sequence ("original nucleic acid sequence") and have the same activity as the original nucleic acid sequences, and also, in particular, natural or artificial mutations of these nucleic acid sequences.

The nucleic acid sequences which can be used for encoding the enzymes having citronellal-isopulegol cyclase activity that can be used in the method according to the invention are likewise described exhaustively in international application PCT/EP2010/057696, hereby incorporated by reference.

With particular preference the method according to the invention is carried out in the presence of an enzyme, the enzyme being encoded by a nucleic acid sequence according to SEQ ID NO: 1 or a functional equivalent thereof, the nucleic acid sequence being part of a gene construct or vector. Such gene constructs or vectors are described exhaustively in international application PCT/EP2010/057696 on pages 16 to 20, hereby incorporated by reference.

With very particular preference the method according to the invention is carried out in the presence of an enzyme, where the enzyme is encoded by a nucleic acid sequence according to SEQ ID NO: 1 or a functional equivalent thereof, the nucleic acid sequence being part of a gene construct or vector which are present in a host cell.

The host cell which comprises a gene construct or a vector in which the nucleic acid sequence is present that encodes the enzyme having the citronellal-isopulegol cyclase activity is also referred to as a transgenic organism. The production of such transgenic organisms is known in principle and is discussed, for example, in international application PCT/EP2010/057696 on page 20, hereby incorporated by reference.

Transgenic organisms selected are preferably cells from the group consisting of bacteria, cyanobacteria, fungi and yeasts. The cell is preferably selected from fungi of the genus *Pichia* or bacteria of the genera *Escherichia*, *Corynebacterium*, *Ralstonia*, *Clostridium*, *Pseudomonas*, *Bacillus*, *Zymomonas*, *Rhodobacter*, *Streptomyces*, *Burkholderia*, *Lactobacillus* or *Lactococcus*. With particular preference the cell is selected from bacteria of the species *Escherichia coli*, *Pseudomonas putida*, *Burkholderia glumae*, *Streptomyces lividans*, *Streptomyces coelicolor* or *Zymomonas mobilis*.

A preferred method according to the invention is that wherein the enzyme having the activity of a citronellal-isopulegol cyclase is encoded by a gene which has been isolated from a microorganism selected from the group of microorganisms consisting of *Zymomonas mobilis*, *Methylococcus capsulatus*, *Rhodospseudomonas palustris*, *Bradyrhizobium japonicum*, *Frankia spec.* and *Streptomyces coelicolor*. With particular preference the gene in question has been isolated from *Zymomonas mobilis*.

Preferred furthermore is a method according to the invention wherein the enzyme having the activity of a citronellal-isopulegol cyclase has been produced by a microorganism which overproduces the enzyme having the activity of a citronellal-isopulegol cyclase and which has been selected from the group of microorganisms consisting of the genera *Escherichia*, *Corynebacterium*, *Ralstonia*, *Clostridium*, *Pseudomonas*, *Bacillus*, *Zymomonas*, *Rhodobacter*, *Streptomyces*, *Burkholderia*, *Lactobacillus* and *Lactococcus*.

A particularly preferred method according to the invention is that wherein the enzyme having the activity of a citronellal-isopulegol cyclase has been produced by transgenic microorganisms of the species *Escherichia coli*, *Pseudomonas putida*, *Burkholderia glumae*, *Corynebacterium glutamicum*, *Saccharomyces cerevisiae*, *Pichia pastoris*, *Streptomyces lividans*, *Streptomyces coelicolor*, *Bacillus subtilis* or *Zymomonas mobilis* which overproduce the enzyme having the activity of a citronellal-isopulegol cyclase.

The above-described further embodiments for carrying out the biocatalytic method according to the invention for cyclizing terpenes apply correspondingly in respect of the production of isopulegol.

A further subject of the present invention is also the use of an enzyme having the activity of a citronellal-isopulegol cyclase for the biocatalytic conversion of citronellal to isopulegol.

Preference is given to the use of an enzyme having the activity of a citronellal-isopulegol cyclase for the biocatalytic conversion of citronellal to isopulegol, wherein the enzyme possesses a polypeptide sequence which either

- a) is SEQ ID NO: 2, or
- b) in which up to 25% of the amino acid residues are altered relative to SEQ ID NO: 2 by deletion, insertion, substitution or a combination thereof, and which still has at least 50% of the enzymatic activity of SEQ ID NO: 2.

Also preferred is the use of an enzyme having the activity of a citronellal-isopulegol cyclase for the biocatalytic conversion of citronellal to isopulegol, wherein the enzyme is encoded by a nucleic acid sequence according to SEQ ID NO: 1 or a functional equivalent thereof.

A further subject of the present invention is also the use of a gene construct or vector comprising a nucleic acid sequence according to SEQ ID NO: 1 or a functional equivalent thereof which encode a polypeptide having the activity of a citronellal-isopulegol cyclase which serves the biocatalytic conversion of citronellal to isopulegol in a method of production of isopulegol by cyclization of citronellal.

Likewise a further subject of the present invention is the use, as well, of a host cell which comprises a gene construct or a vector comprising a nucleic acid sequence according to SEQ ID NO: 1 or a functional equivalent thereof for producing an enzyme having the activity of a citronellal-isopulegol cyclase for the biocatalytic conversion of citronellal to isopulegol.

The method described above opens up for the first time the possibility of cyclizing citronellal to isopulegol by means of an enzyme.

#### 8. Methods of Production of Menthol

The isopulegol prepared inventively can be converted into menthol by catalytic hydrogenation in a conventional way. Suitable for this purpose, as well as conventional hydrogenation processes, is, in particular, a catalytic method, as described in WO 2009/013192.

The method according to the invention is implemented in particular using catalysts comprising

- 45% to 55% by weight of oxygen-containing compounds of nickel, calculated as NiO,
- 25% to 35% by weight of oxygen-containing compounds of zirconium, calculated as ZrO<sub>2</sub>,
- 5% to 20% by weight of oxygen-containing compounds of copper, calculated as CuO,
- 1% to 3% by weight of oxygen-containing compounds of molybdenum, calculated as MoO<sub>3</sub>, and
- 0% to 5% by weight of further components,

the figures in % by weight adding up to 100% by weight and relating to the dry, unreduced catalyst.

One particularly preferred catalyst is composed of 49% to 53% by weight of NiO, 15% to 19% by weight of CuO, 28% to 32% by weight of ZrO<sub>2</sub> and 1% to 2% by weight of MoO<sub>3</sub> and also, optionally, 0% to 3% by weight of further components such as graphite, for example, the respectively selected weight fractions of the individual components being based on the dry, unreduced catalyst and adding up to 100% by weight. Catalysts of this kind are known and can be produced for example as described in EP 0 696 572 or in WO 2009/013192.

In general the catalysts are used preferably in the form of unsupported catalyst. The term "unsupported catalyst" refers to a catalyst which in contrast to a supported catalyst is composed only of catalytically active material. Unsupported catalysts can be used by introducing the catalytically active material, ground to a powder, into the reaction vessel, or by disposing the catalytically active material in the reactor after grinding, mixing with shaping aids, shaping and heat-treating in the form of shaped catalyst bodies—for example, as spheres, cylinders, tablets, rings, coils, strands and the like.

In the context of one preferred embodiment of the hydrogenation method according to the invention, the selected heterogeneous catalyst is employed in the form of a fixed-bed catalyst.

To implement the method according to the invention, the isopulegol starting material as described above is contacted with hydrogen and with the selected catalyst. The hydrogen here may be used in undiluted form, typically in a purity of about 99.9% by volume, or in diluted form, i.e. in the form of mixtures with inert gases such as nitrogen or argon, for example. It is preferred to use hydrogen in undiluted form. The reaction can be carried out successfully without adding solvent or in the presence of organic solvents which are inert under the reaction conditions, such as, for example, methanol, ethanol, isopropanol, hexane, heptane, cyclohexane and the like. It is preferred to carry out the reaction without adding solvent.

The hydrogenation of isopulegol in accordance with the invention can be carried out under a hydrogen pressure (absolute) in the range from 1 to 200 bar, such as from 2 or 3 to 200 bar, in particular from 4 or 5 to 150 bar, such as from 5 to 100 bar, or in the range from 5 to 50 bar. As a reaction temperature for implementing the hydrogenation according to the invention, a temperature is selected, advantageously, that is in the range from 20 to 150° C., such as from 40 to 130° C., or from 60 to 110° C. and more particularly from 70 to 100° C.

The practical approach to the implementation is generally to supply the isopulegol for conversion to the catalyst, which is typically located in a fixed bed reactor heated, in particular, from the outside, such as a tube reactor, autoclave or tube-bundle reactor, for example, at the desired reaction temperature and under the desired pressure. The velocity over the catalyst in this case is generally 0.1 to 1.0, such as 0.1 to 0.6 or 0.2 to 0.4, kg of isopulegol per kg of catalyst per hour. In this context it may be useful to heat the isopulegol that is to be used, even before it is supplied to the reaction vessel or to the reactor, this heating being preferably to reaction temperature.

The reactor can be operated either in liquid phase mode or in trickle mode—that is, the starting materials may be passed through the reactor either from bottom to top or from top to bottom. The hydrogenation method of the invention can be

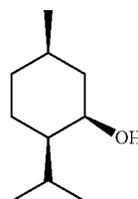
carried out either batchwise or continuously. In both cases, unreacted starting material can be circulated together with the hydrogen.

The hydrogenation according to the invention may also be carried out in stages in a cascade of two or more reactors, i.e. 2 to in general 4, such as 2 or 3, for example, reactors connected in series, preferably fixed bed reactors. In this case, in the first reactor, typically referred to as the main reactor, the main conversion of the reaction is achieved under the reaction conditions described above, and the crude product obtained is passed to a second reactor, typically referred to as secondary reactor, in which the as yet unreacted starting material is at least largely converted inventively into L-menthol. The reaction conditions here may be selected, independently of one another, preferably in the ranges stated above.

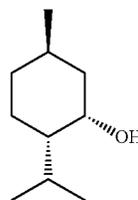
The method of the invention can be carried out batchwise, semibatchwise or continuously. It is preferred to carry out the method continuously, more particularly entirely continuously, in which case the starting materials are introduced continuously into the reactor and the resulting reaction mixture or reaction product is discharged continuously from the reactor. It has further proven advantageous, in view of the position of the melting point of the reaction product according to the invention, namely menthol, especially L-menthol, to provide for heating of the transport lines used.

The method of the invention allows menthol to be produced by catalytic hydrogenation of isopulegol, with typically only a minor degree of formation of unwanted diastereomers of menthol. Accordingly, when using isopulegol with a corresponding purity, the method of the invention yields menthol of the formula (III) in a chemical purity of 97% by weight or more, preferably of 98% to 100% by weight, more preferably of 98.5% to 99.9% by weight, very preferably at least 99% to 99.9% by weight. The term "chemical purity" here also encompasses the diastereomeric purity of the resulting menthol in relation to the diastereomers neoisomenthol of formula (IIIa), neomenthol of formula (IIIb) and isomenthol of formula (IIIc). Accordingly, in the context, the method according to the invention preferably yields menthol having a diastereomeric purity of 97% by weight or more, preferably of 98% to 100% by weight, more preferably of 98.5% to 99.9% by weight and very preferably of at least 99% to 99.9% by weight.

(IIIa)

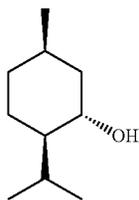


(IIIb)



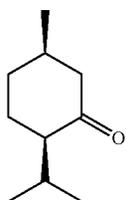
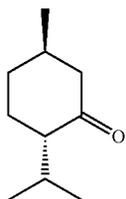
49

-continued



Where isopulegol is used in optically active form—preferably, in accordance with the invention, mixtures comprising predominantly the L-isopulegol enantiomer—the method product according to the invention that is obtained is generally menthol in optically active form, preferably in the form of (-)- or L-menthol. The hydrogenation according to the invention proceeds generally largely without notable racemization of the material used. Accordingly, according to the enantiomeric excess of the optically active isopulegol used, optically active menthol, preferably L-menthol when using L-isopulegol, is obtained as the product, with an enantiomeric excess (ee) of 80% ee or more, preferably of 85% or 90% ee or more, more preferably of 95% to 100% ee, more preferably of 96% to 99.9% ee, very preferably of 97% to 99.8% ee, even more preferably of 98% to 99.7% ee, and with more particular preference of 98.5% to 99.6% ee.

The menthol obtained according to the invention is notable, furthermore, for a particularly low level of the unwanted by-products menthone of formula (IIIc) and isomenthone of formula (IIIe) and neoisomenthol of formula (IIIa).



These by-products are obtained generally, in the context of the method according to the invention, only in a proportion, relative to the amount of menthol obtained, of up to 0.5% by weight, preferably 0.4% by weight, more preferably 0.3% by weight, more particularly 0.2% by weight, and very preferably 0.1% to 0% by weight.

9. Examples of Substrates which can be Used for Enzymatic or Biocatalytic Conversions According to the Invention:

The enzymes and microorganisms described herein are especially suitable for converting compounds of the general formula IV above. Non-limiting examples thereof are summarized in table A below, which gives the structural formula and the chemical name.

50

TABLE A

(IIIc)		Further substrates	
		Formula	
5			
10		(IV)	Name
15			Citral
20			Neral
25			Nerol
30			Nerylacetone
35	(IIIc)		Geranial
40			Geraniol
45	(IIIe)		
50			
55			
60			
65			

51

TABLE A-continued

Further substrates	
Formula	
(IV)	Name

	Geranylic acid
	cis-Geranylic acid
	Geranylacetone
	Farnesol
	Farnesylacetone

52

TABLE A-continued

Further substrates	
Formula	
(IV)	Name

	Homofarnesylol
	Homofarnesol
	Trimethyltrideca-tetraene
	Melonal
	Nonadienal

53

TABLE A-continued

Further substrates	
Formula	Name
<p>(IV)</p>	
	Citronellol
	$\beta$ -Citronellene
	Citronellic acid
	Hydroxycitronellal
	Heptanal
	Linalool

54

TABLE A-continued

Further substrates	
Formula	Name
<p>(IV)</p>	
	Farnesene ( $\beta$ )
	Myrcene
	Myrcenol
	Dihydromyrcenol
	Lavandulol
	Nerolidol

TABLE A-continued

Further substrates	
Formula	Name
<p>(IV)</p>	
	(E)- $\beta$ -Ocimene (4 isomers present)
	Tagetone
	Solanone
	2,6,10-Trimethyl- 9-undecanal

The reaction products produced in the conversion of these substrates can be detected and quantified in a conventional way using standard analytical methods, such as chromatography, HPLC, gas chromatography, mass spectrometry, GC/MS or MALDI-TOF, and combinations thereof.

If nonimmobilized organisms or enzymes are used for the method according to the invention, preferably these are separated prior to extraction, for example by filtration or centrifugation.

The method according to the invention can be operated batchwise, semi-batchwise or continuously.

#### EXPERIMENTAL SECTION

In the absence of special information in the examples below, the general information below is taken to apply.

##### A. General Information

All materials and microorganisms used are commercially available products.

Unless stated otherwise, the cloning and expression of recombinant proteins is carried out by standard methods, as described for example in Sambrook, J., Fritsch, E. F. and Maniatis, T., *Molecular Cloning: A Laboratory Manual*, 2nd edition, Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.

##### a) Bacterial Strains, Plasmids and Growing Conditions

All experiments were carried out with *E. coli*. The SHC proteins were expressed in *E. coli* BL21 (DE3) pLysS or *E. coli* Rosetta pLysRAR62, comprising pET16b constructs with the respective *she* gene, by growing in Luria-Bertani medium, supplemented with ampicillin (100  $\mu$ g/ml), chloramphenicol (34  $\mu$ g/ml), and 0.5 mM isopropylthio- $\beta$ -D-galactoside at OD<sub>600</sub> of 0.4 and additional growth for 4 hours at 30° C.

##### b) Vector Constructs

The respective squalene-hopene cyclase gene (e.g. *Zymomonas mobilis* ZMO1548 [NC\_006526.2, region: 1578816 . . . 1580993]) was PCR-amplified from chromosomal DNA, using corresponding primer pairs (e.g. ZMO1548-fwd (5'-gcgctgttcatatgggtattgaca-3') (SEQ. ID. NO: 327) and ZMO1548-rev (5'-gcgcttaccctggatcctcgaaaat-3') (SEQ. ID. NO: 328)). The restriction enzyme digested (e.g. with NdeI/BamHI) PCR product was cloned into pET16b, (obtaining e.g.) pET1584. The constructs were verified by DNA sequencing and transformed into *E. coli* XL1-blue.

The *she*-gene from other microorganisms (e.g. from *A. acidocaldarius*) was cloned similarly.

All plasmids were transformed individually into *E. coli* BL21 (DE3) pLysS or *E. coli* Rosetta pLys-RAR62.

##### c) Cyclization Assay with Various Substrates (Standard Conditions)

Recombinant *E. coli* cells were suspended in 20 mM Tris-HCl pH 8.0 (3 ml per g moist cells).

The cyclization mixture contained 250  $\mu$ l of cell suspension, 50  $\mu$ l of 1 M citrate buffer (pH 4.5), 20 mM (final concentration) of substrate and water to 500  $\mu$ l. In the cyclization of squalene, 1% (v/v) Triton-X100 was added. For the homofarnesol cyclization, *E. coli* cells (6 g moist cells) were suspended in solubilization buffer (50 mM phosphate, 10 mM MgCl<sub>2</sub> (pH 6.5; total volume: 25 ml). The cells were lysed at 1500 bar using a Manton-Gaulin homogenizer. Insoluble cellular debris was centrifuged off (15 min at 4° C. and 7150 g). The cyclization mixture contained 1 ml raw cell extract and 20 mM homofarnesol in 1.25 ml buffer (50 mM potassium phosphate, 45 mM MgCl<sub>2</sub> (pH 6.5)). The reaction mixture was stirred at 30° C. with a magnetic stirrer. The reaction was stopped by extraction with heptane. The organic phase was analyzed by gas chromatography. Controls were carried out with *E. coli* cells bearing an empty vector and with heat-inactivated SHC-expressing cells. Formation of cyclization products was never observed with the controls (data not shown).

##### d) Gas Chromatography

Terpenoids were analyzed qualitatively and quantitatively by gas chromatography using an Agilent 7890A gas chromatograph, equipped with a DB-5 column (20 m $\times$ 0.1 mm $\times$ 0.1  $\mu$ m) and an ionization detector. 3  $\mu$ l of the solvent extract was applied on the column (split ratio 1:5, helium flow rate 0.25 or 0.5 ml/min, injector temperature 250° C.).

To separate linear and cyclic monoterpenoids, the initial furnace temperature (60° C.) was raised to 130° C. at 40° C./min, at 2° C./min to 150° C. and threat 40° C./min to 200° C. The retention times of the terpenoids were as follows: (R, S)-citronellal (7.55 min), isopulegol (7.70 min), neo-isopulegol (7.90 min), iso-isopulegol (8.10 min), neoiso-isopulegol (8.25 min), 1-decanol (9.91 min).

For the detection of triterpenes, the injector temperature was set at 300° C. The furnace temperature was initially 60°

57

C., and was increased at 40° C./min to 220° C. and then at 6° C./min to 310° C. and held constant there for 10 min. Squalene and hopene eluted after 19.2 min and 26.9 min respectively.

Homofarnesol and ambroxan were analyzed on a 10 m Optima 1 column (Macherey&Nagel, Düren, Germany). The initial furnace temperature (100° C.) was increased at 5° C./min to 200° C. and held at this temperature for 5 min. Then it was increased at 30° C./min to 320° C. An analysis took 40 min. The retention times were as follows: homo-

farnesol (10.8 min), ambroxan (9.9 min). As an alternative, a Shimadzu GC-MS QP 2010 system with an FS Supreme 5 column (30 m×0.25 mm×0.25 μm) was used for coupled GC/MS analysis (split ratio 1:20; 3 min 120° C., increase to 135° C. at 2° C./min and further increased to 365° C. at 10° C./min, followed by cooling to 300° C. at 70° C./min). The GC-MS data were analyzed using LabSolutions GCsolutions Postrun software. It should be noted that the substrates citronellal racemate, (R)-citronellal and (S)-citronellal always contain small amounts of isopulegol and neo-isopulegol as impurities. The GC surface values for these linear terpenoids were established as 100%. The surface values for the isopulegol isomers in the product were corrected by the amount of isopulegol isomer that was already present in the substrate. The standard deviation was calculated on the basis of 24 individual tests using two separately grown *E. coli* cultures.

## B. EXAMPLES

### Example 1

#### Production of Mutants of the F486X Type of the Squalene-hopene Cyclases by Rational Protein Design Using Quick-change Mutagenesis

The mutants of various squalene-hopene cyclases were incorporated by means of "quick-change" mutagenesis into

58

the corresponding gene. The procedure based on the manufacturer's information (Agilent Technologies, Waldbronn) was largely followed. First, a PCR was carried out:

PCR charge: 1.8 μl DMSO  
2 μl dNTPs (each 2.5 mM)  
1.5 μl forward primer (10 μmol/μl)  
1.5 μl reverse primer (10 μmol/μl)  
1 μl templates (1 μg/μl; recombinant plasmid bearing SHC gene, for example pETZmSHC\_1)  
0.2 μl Prime-Star Polymerase (Takara, 2.5 Units/μl)  
6 μl 5× buffer  
16 μl H<sub>2</sub>O

PCR Program:

- (1) 95° C. 3 minutes
  - (2) 95° C. 45 seconds
  - (3) 53° C. 1 minute
  - (4) 68° C. 17 minutes
- 5× repetition of steps (2), (3) and (4)

After the PCR, 10 μl of the charge was digested with the restriction enzyme DpnI for at least 1 hour at 37° C. Then transformation into *E. coli* XL1-blue cells was carried out. After DNA sequencing, transformation into the expression strain e.g. *E. coli* Rosetta pLysRAR62 took place. The gene can also be modified similarly in other expression plasmids.

The following primers were used for the quick-change PCR. The respective exchange is shown printed in bold in the primer names. The genes that are modified by the respective primers are indicated with italics in the primer names; there is the following correspondence:

Primer name	Sequence	SEQ ID NO
<i>ZmSHC_1F486I1e</i> for	GTTATTATCCTTATCGATGGCTCCCAACCG	329
<i>ZmSHC_1F486I1e</i> rev	GGTTGGGGAGCCATCGATAAGGATAATAACAG	330
<i>ZmSHC_1F486Met</i> for	GTTATTATCCTTATCCATGGCTCCCAACCG	331
<i>ZmSHC_1F486Met</i> rev	GGTTGGGGAGCCATCGATAAGGATAATAACAG	332
<i>ZmSHC_1F486Thr</i> for	GTTATTATCCTTATCGGTGGCTCCCAACCG	333
<i>ZmSHC_1F486Thr</i> rev	GGTTGGGGAGCCACCGATAAGGATAATAACAG	334
<i>ZmSHC_1F486G1n</i> for	GTTATTATCCTTATCCTGGGCTCCCAACCG	335
<i>ZmSHC_1F486G1n</i> rev	GGTTGGGGAGCCAGGATAAGGATAATAACAG	336
<i>ZmSHC_1F486Asn</i> for	GTTATTATCCTTATCGTTGGCTCCCAACCG	337
<i>ZmSHC_1F486Asn</i> rev	GGTTGGGGAGCCAACGATAAGGATAATAACAG	338
<i>ZmSHC_1F486Lys</i> for	GTTATTATCCTTATCTTTGGCTCCCAACCG	339
<i>ZmSHC_1F486Lys</i> rev	GGTTGGGGAGCCAAGATAAGGATAATAACAG	340
<i>ZmSHC_1F486Asp</i> for	GTTATTATCCTTATCATCGGCTCCCAACCG	341
<i>ZmSHC_1F486Asp</i> rev	GGTTGGGGAGCCGATGATAAGGATAATAACAG	342
<i>ZmSHC_1F486Glu</i> for	GTTATTATCCTTATCTTCGGCTCCCAACCG	343
<i>ZmSHC_1F486Glu</i> rev	GGTTGGGGAGCCGAAGATAAGGATAATAACAG	344
<i>ZmSHC_1F486Trp</i> for	GTTATTATCCTTATCCCAGGCTCCCAACCG	345

- continued

Primer name	Sequence	SEQ ID NO
<i>ZmSHC_1F486Trp</i> rev	GGTTGGGGAGCCTGGGATAAGGATAATAACAG	346
<i>ZmSHC_1F486Arg</i> for	GTTATTATCCTTATCACGGGCTCCCCAACCG	347
<i>ZmSHC_1F486Arg</i> rev	GGTTGGGGAGCCCGTGATAAGGATAATAACAG	348
<i>ZmSHC_1F486Cys</i> for	GTTATTATCCTTATCGCAGGCTCCCCAACCG	349
<i>ZmSHC_1F486Cys</i> rev	GGTTGGGGAGCCTGCGATAAGGATAATAACAG	350
<i>ZmSHC_1F486G</i> for	GTTATTATCCTTATCACGGGCTCCCCAACCG	351
<i>ZmSHC_1F486G</i> rev	GGTTGGGGAGCCGGTGATAAGGATAATAACAG	352
<i>ZmSHC_1F486S</i> for	GTTATTATCCTTATCGCTGGCTCCCCAACCG	353
<i>ZmSHC_1F486S</i> rev	GGTTGGGGAGCCAGCGATAAGGATAATAACAG	354
<i>ZmSHC_1F486P</i> for	GTTATTATCCTTATCCGGGGCTCCCCAACCG	355
<i>ZmSHC_1F486P</i> rev	GGTTGGGGAGCCCGGATAAGGATAATAACAG	356
<i>ZmSHC_1F486H</i> for	GTTATTATCCTTATCATGGGCTCCCCAACCG	357
<i>ZmSHC_1F486H</i> rev	GGTTGGGGAGCCCATGATAAGGATAATAACAG	358
<i>ZmSHC_1F486L</i> for	GTTATTATCCTTATCCAGGGCTCCCCAACCG	359
<i>ZmSHC_1F486L</i> rev	GGTTGGGGAGCCCTGGGATAAGGATAATAACAG	360
<i>ZmSHC_1F486V</i> for	GTTATTATCCTTATCAACGGCTCCCCAACCG	361
<i>ZmSHC_1F486V</i> rev	GGTTGGGGAGCCGTGATAAGGATAATAACAG	362
<i>ZmSHC_1F486A</i> for	GTTATTATCCTTATCCGGGCTCCCCAACCG	363
<i>ZmSHC_1F486A</i> rev	GGTTGGGGAGCCCGGATAAGGATAATAACAG	364
<i>ZmSHC_1F486Y</i> for	GTTATTATCCTTATCATAGGCTCCCCAACCG	365
<i>ZmSHC_1F486Y</i> rev	GGTTGGGGAGCCTATGATAAGGATAATAACAG	366
<i>ZmSHC_1Y702C</i> for	GCCGATAAAAATCGCAACGCAGCATAAACG	367
<i>ZmSHC_1Y702C</i> rev	CGTTTATGCTGCGTTGCGATTTTATCGGC	368
<i>ZmSHC_1Y702F</i> for	GCCGATAAAAATCTTACGCAGCATAAACG	369
<i>ZmSHC_1Y702F</i> rev	CGTTTATGCTGCGTAAAGATTTTATCGGC	370
<i>ZmSHC_1Y702A</i> for	GCCGATAAAAATCCGCACGCAGCATAAACG	371
<i>ZmSHC_1Y702A</i> rev	CGTTTATGCTGCGTGCGGATTTTATCGGC	372
<i>ZmSHC_1Y702S</i> for	GCCGATAAAAATCGCTACGCAGCATAAACG	373
<i>ZmSHC_1Y702S</i> rev	CGTTTATGCTGCGTAGCGATTTTATCGGC	374
<i>ZmSHC_1Y561A</i> for	GAACCGCACCCGGTGCCATAGATCGCATTAAACG	375
<i>ZmSHC_1Y561A</i> rev	GGTTTGGTCGTTGGGGCGTTAATGCGATCTATGG	376
<i>ZmSHC_1Y705A</i> for	CCATAATCGGAAGAATGCCCGCGCAAAATC	377
<i>ZmSHC_1Y705A</i> rev	CTGCGTTATGATTTTTCGCGGCAATTCTTC	378
<i>ZmSHC_2F486C</i> for	GGCGGTTGGGGCGCTTGCATGCCAATAACAG	379
<i>ZmSHC_2F486C</i> rev	CTGTTATTGGCATCGCAAGCGCCCCAACCGCC	380
<i>ApF486C</i> rev	CATTATCTTTATCGCATGCACCCCAACCACC	381
<i>ApF486C</i> for	GGTGGTTGGGTGCATGCGATAAAGATAATG	382
<i>BjF486C</i> for	CGGCTGGGGCGCGTGCGATAAAGATAAC	383

- continued

Primer name	Sequence	SEQ ID NO
<i>Bj</i> F486Crev	GTTATCTTTATCGCACGCGCCCCAGCCG	384
<i>Sc</i> F486Cfor	CGGCGCCTGGGGCGCCTGCGACGTCGACAAC	385
<i>Sc</i> F486Crev	GTTGTCGACGTCGCAGGCGCCCCAGGCGCCG	386

*Zm*SHC\_1 SEQ ID NO: 2;*Zm*SHC\_2 SEQ ID NO: 6;*Ap* SEQ ID NO: 4;*Bj* SEQ ID NO: 5 and*Sc* SEQ ID NO: 3.

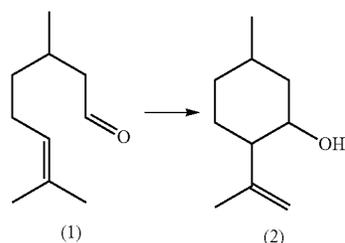
## Example 2

Activity Tests with Mutants of Squalene-Hopene Cyclase-1 (SHC-1) from *Zymomonas Mobilis*

The influence of various single mutations, produced according to example 1, in the sequence position corresponding to F486, on the cyclase activity was determined for various substrates.

## a) Citronellal

After the general detection of a slight cyclization activity of the squalene-hopene cyclase-1 from *Zymomonas mobilis* (SEQ ID NO:2) with respect to citronellal, the turnover rate was greatly improved by rational protein design. Exchange of the phenylalanine residue F486 for alanine led in preliminary tests (cf. FIG. 2) to a greatly increased production of isopulegol (2) starting from citronellal (1).



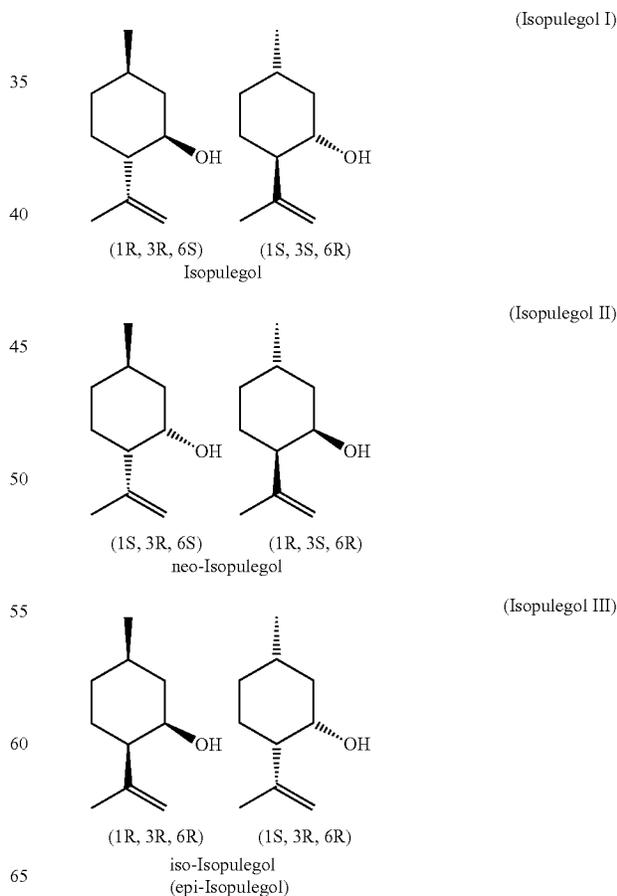
The increased activity of the SHC\_1-F486A mutant was then investigated in more detail. In addition to a far better conversion of the citronellal substrate, it was also found that this prefers the R(+) isomer as substrate and compared with the WT it is also converted in a much shorter time (cf. FIG. 2). Whereas with the WT enzyme the reaction with R(+)-citronellal is not measurable until after quite long incubation, the F486A mutant shows high conversions, in particular at the start of the reaction. This effect is not observed with S(-)-citronellal as substrate. It is notable that the F486A mutant only forms isopulegol I and II, whatever the stereo-configuration of the substrate. The WT, in contrast, is dependent on the stereoconfiguration of the substrate and forms, as well as isopulegol I, mainly isopulegol II from R(+)-citronellal and almost exclusively isopulegol III from S(-)-citronellal.

Based on these results, in further experiments the importance of the amino acid residues at position 486 was investigated more closely. For this, by means of mutagenesis, the phenylalanine residue was exchanged against each further amino acid and the activity of the various muteins was tested with citronellal as substrate (for sequences see FIGS. 1a and b). It was found that some amino acids at this position not

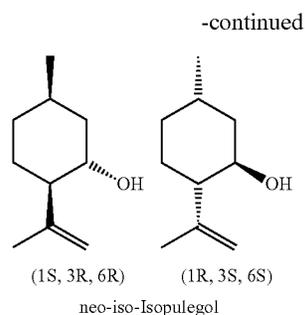
only improve the conversion of citronellal by the enzyme, but additionally lead to higher product specificity in the reaction, so that fewer isomers of isopulegol are produced (see FIG. 3).

Exchange for arginine, proline and lysine leads to a loss in activity with respect to citronellal. The amounts of product determined also occur, in the same distribution, as contamination in the negative control ('K' see FIG. 3). The highest activity was observed after exchange for valine, threonine, cysteine, isoleucine and alanine. Overall, the altered product spectrum of some muteins is notable. Not all show the formation of three isopulegol peaks as the wild type as well as the quantitative distribution differs.

There are altogether  $2^3$  isopulegol isomers:



63



(Isopulegol IV)

Until now, the main product (isopulegol I) has been assigned to the enantiomeric pair (1R,3R,6S)-isopulegol or (1S,3S,6R)-isopulegol.

The highest yield of isopulegol with the least by-products (consisting of further isomers) accompanied by high enzyme activity is displayed by the Zm-SHC-1 F486C mutant.

#### b) Squalene

Clear changes in activity after mutation at position F486 are also seen with squalene as substrate. Interestingly, in this case the exchange of phenylalanine for tyrosine produces almost a doubling of the conversion (see FIG. 4).

### Example 3

#### Activity Tests with Mutants of other Squalene-hopene Cyclases

The influence of various single mutations, produced according to example 1, in the sequence position corresponding to F486 on the cyclase activity of various other SHCs was determined for various citronellal substrates (in each case 20 mM overnight incubation):

The mutants are as follows:

Ap-SHC: F481C,  
Bj-SHC: F447C,  
Sc-SHC: F449C,  
Zm SHC-2: F438C

The phenylalanine residues are located in positions that are analogous to the F486 of Zm-SHC-1 (SEQ ID NO:2).

The results can be seen in FIG. 5 (citronellal racemate as substrate), FIG. 6 (R(+)-citronellal as substrate), and FIG. 7 (S(-)-citronellal as substrate). The control was a charge without active biocatalyst.

It can be seen that the wild-type enzymes, through mutation at the stated position corresponding to F486 (of Zm SHC-1), can now cyclize citronellal to isopulegol and moreover convert the R(+) form with increased selectivity compared with the S(-) form.

### Example 4

#### Conversion of Compounds of Formula IV

These substances were converted under conditions corresponding to those employed for the conversion of citronellal as described above.

### Example 5

#### Isolation and Characterization of the Squalene-hopene Cyclase from *Zymomonas Mobilis* (Zm-SHC)

International application PCT/EP2010/057696, hereby incorporated by reference, describes how, using specific

64

oligonucleotides, the Zm-SHC gene from the genomic DNA of *Zymomonas mobilis* was amplified and expressed in *Escherichia coli*.

#### a) Material and Methods:

5 Addressed below are only materials and methods not mentioned in this form in international application PCT/EP2010/057696.

#### b) Strains, Plasmids and Culture Conditions:

10 The *E. coli* strain DH5 $\alpha$ , the *E. coli* strain BL21 (DE3) pLysS (Novagen) and the *E. coli* Rosetta strain were used. The plasmid pET16b (Novagen) was used for cloning. For the overexpression of the SHC, moreover, the plasmid pLysRAR62 was additionally transformed for the adaptation of the codon usage to *E. coli*. Furthermore, the plasmid pDHE+ZmSHC-1 from *E. coli* Lu15568 was used (international application PCT/EP2010/057696). The strains were grown using LB medium at 30° C.

#### c) Chemicals:

20 Squalene, (+/-)-citronellal, (+)-R-citronellal and (-)-S-citronellal were purchased from Sigma (Sigma-Aldrich Chemie GmbH, Munich). Restriction enzymes, T4 ligase, and DNA polymerase came from New England Biolabs (New England Biolabs GmbH, Frankfurt).

#### d) Isolation of DNA and Transformation:

25 Plasmids were isolated from *E. coli* using the Qiaprep Spin Miniprep Kits from Qiagen (Qiagen, GmbH, Hilden). For gel extractions or PCR purifications, the Qiaquick Gel Extraction Kit from Qiagen was used. All of the *E. coli* strains used were transformed using the CaCl<sub>2</sub> method.

#### e) PCR and Sequencing:

30 The DNA from *Zymomonas mobilis* subsp. *mobilis* CP4 was provided by Prof. Sprenger (Institute of Microbiology, University of Stuttgart). The PCR was carried out using Prime Star Polymerase. The following primers were used for synthesizing the squalene-hopene cyclase gene from *Zymomonas mobilis*:

```

SHC_1:
SHC-f or TATGCATATGGGTATTGACAGAAT
(SAQ ID NO: 387)
SHC-rev CCGGATCCTCAATTATTCAATCAATC
(SAQ ID NO: 388)

```

45 The correctness of the cloned genes was verified by means of sequencing by the company GATC Biotech. Sequence analyses were carried out using the program Clone Manager 7.0. After restriction of the corresponding amplicates, they were cloned in-frame into the pET16b vector using N-terminally encoded His-tag. The plasmids were subsequently transformed first in *E. coli* DH5 $\alpha$  and thereafter in *E. coli* BL21 (DE3)pLysS and *E. coli* Rosetta. For better expression, the plasmid pLysRAR62 was transformed into the *E. coli* Rosetta strains in addition to the pET16b constructs. Corresponding clonings with empty vectors were carried out in parallel. In addition, the plasmid pDHE+ZmSHC\_1 (corresponding to SHC\_1 with codon usage adapted to *E. coli*) was transformed in *E. coli* BL21 (DE3) pLysS.

#### f) Expression and Cell Digestion:

60 The corresponding *E. coli* BL21 (DE3) pLysS and *E. coli* Rosetta transformants were cultured in LB medium with ampicillin and chloramphenicol (100  $\mu$ g/ml and 32  $\mu$ g/ml, respectively) at 30° C. The synthesis of the squalene-hopene cyclases was induced by addition of 0.5-1 mM IPTG or 0.1% rhamnose (when using the pDHE derivatives) with an OD<sub>600</sub> of 0.4-0.6. The cells were allowed to grow further for

65

4-6 hours, and subsequently harvested. This was done by centrifuging off the cells and taking them up in 5 ml/g wet weight of 25 mM Tris/HCl with 40% glycerol. If the cells were not used further immediately, they were stored at -20° C. For digestion of the cells, they were each subjected 2x to a French Press and used, either directly or following removal of the cell debris by centrifugation, for the activity assays. Alternatively, cell digestion took place using ultrasound. Following centrifugation, the SHC proteins were subsequently dissolved with solubilization buffer (50 mM Tris/HCl pH 8, 10 mM MgCl<sub>2</sub>, 1% Triton X-100) to remove the cell debris, and hence partially enriched.

g) Activity Assays:

Each batch for determining the activity of the squalene-hopene cyclases had a final volume of 1 ml. This was made up of 600 µl of cells digested by French Press (alternatively 800 µl after solubilization from the cell membrane), 100 mM Na citrate buffer with different pH levels (pH 4.0 to pH 8.0 were used for testing) and 10 mM substrate solution [(+/-) citronellal, (+)-R-citronellal and (-)-S-citronellal]. In addition to the substrate and H<sub>2</sub>O, the substrate solution also comprised Triton X-100, which was present in each of the activity batches at a concentration of 0.2%.

The batches were incubated with shaking for 6 hours to 24 hours at temperatures of 22° C., 30° C. and 37° C. The substrate and possible products were extracted with one volume of chloroform or hexane/propanol in a ratio of 2:3. The extract was used directly for analysis by gas chromatography.

h) GC Measurements:

The gas-chromatographic measurements took place on an Agilent 7890A gas chromatograph with flame ionization detector. The column used was a DB-5 (Agilent Technologies) with a length of 20 m, a diameter of 0.1 mm and 0.25 µM coating. Substances were identified by comparison of the retention times with available standard solutions.

For verification, the samples were analyzed in parallel on a Shimadzu Gas chromatograph with mass spectrometer. Using the column FS Supreme with a length of 30 m, an internal diameter of 0.25 mm and a coating of 0.25 µm, the retention times were again compared with standard solutions, and the respective mass spectra of the substances present were analyzed.

With the aid of a standard, the diastereomer identified below as isopulegol I was assigned to (1R,3R,6S) or (1S,3S,6R) isopulegol, whereas no assignment was possible for the isomers identified as isopulegol II and isopulegol III.

66

i) Results of the Activity Assays:

1. Test 1a: (comparative) (controls i.e. results with boiled-off protein, with empty vector and without protein)

	pH 4.0	pH 4.5	pH 5.0	pH 5.5	pH 6.0	pH 6.5	pH 7.0
Citronellal	85.4	85.4	86.0	85.6	84.4	84.7	85.1
Isopulegol I	10.8	10.8	10.4	10.8	11.7	11.5	11.2
Isopulegol II	3.8	3.8	3.6	3.6	3.9	3.8	3.7
Isopulegol III	0	0	0	0	0	0	0

In the information below concerning the substrate rac-citronellal, take place with the amounts of isopulegol found in the controls having already been deducted.

2. Test 1b: Comparison of the two overexpressed SHC\_1 proteins (from pDHE and pET16b vector and influence of the His-tag on activity at pH 4.5)

	pDHE	pET16b
Citronellal	95.2	95.2
Isopulegol I	0.7	0.8
Isopulegol II	1.7	1.6
Isopulegol III	2.4	2.4

3. Test 1c: pH dependence

	pH 4.0	pH 4.5	pH 5.0	pH 5.5	pH 6.0	pH 6.5	pH 7.0
Citronellal	95.9	94.9	94.7	94.4	95.1	98.7	98.8
Isopulegol I	0.4	0.8	0.8	1.0	1.1	0.8	0.5
Isopulegol II	1.1	2.4	2.1	2.1	1.6	0.5	0.7
Isopulegol III	2.6	1.9	2.4	2.5	2.2	0	0

4. Test 1d: Influence of salts at pH 4.5

	none	BaCl <sub>2</sub>	CaCl <sub>2</sub>	MgCl <sub>2</sub>
Citronellal	94.9	95.2	94.9	95.0
Isopulegol I	0.7	0.8	1.0	0.9
Isopulegol II	2.5	2.4	2.4	2.5
Isopulegol III	1.9	1.6	1.7	1.6

5. Test 1e: Influence of temperature at pH 4.5

	22° C.	30° C.	37° C.
Citronellal	95.3	94.9	95.4
Isopulegol I	0.8	1.0	0.8
Isopulegol II	1.8	2.2	1.6
Isopulegol III	2.1	1.9	2.2

6. Test 2: S(-)-Citronellal as substrate

	pH 4.0	CTRL	pH 4.5	CTRL	pH 5.0	CTRL	pH 5.5	CTRL
Citronellal	90.8	95.5	90.8	95.7	91.7	96.2	92.4	96.2
Isopulegol I	4.9	4.5	4.7	4.3	4.4	3.8	4.1	3.8
Isopulegol II	0	0	0	0	0	0	0	0
Isopulegol III	4.3	0	4.5	0	3.9	0	3.5	0
	pH 6.0	CTRL	pH 6.5	CTRL	pH 7.0	CTRL		
Citronellal	94.1	96.6	96.4	96.5	96.5	96.4		
Isopulegol I	3.8	3.4	3.6	3.5	3.5	3.6		
Isopulegol II	0	0	0	0	0	0		
Isopulegol III	2.1	0	0	0	0	0		

7. Test 3: R-(+)-Citronellal as substrate

	pH 4.0	CTRL	pH 4.5	CTRL	pH 5.0	CTRL	pH 5.5	CTRL
Citronellal	80.0	84.2	78.4	83.8	81.1	85.6	81.7	86.8
Isopulegol I	15.9	15.8	16.0	16.2	14.1	14.4	13.5	13.2
Isopulegol II	4.1	0	5.6	0	4.8	0	4.8	0
Isopulegol III	4.3	0	4.5	0	3.9	0	3.5	0

	pH 6.0	CTRL	pH 6.5	CTRL	pH 7.0	CTRL
Citronellal	81	85.5	80.8	85.8	81.4	86.2
Isopulegol I	14.3	14.5	14.5	14.2	14.0	13.8
Isopulegol II	4.7	0	4.7	0	4.6	0
Isopulegol III	2.1	0	0	0	0	0

15

j) Summary of the Results:

The squalene-hopene cyclase from *Zymomonas mobilis* was prepared recombinantly in *E. coli*. The enzyme is able to convert citronellal to isopulegol.

Here, the two overproduced Zm-SHC-1 proteins, once without and once with N-terminally appended His-tag, showed no differences in their activity under the conditions tested (cf. Test 1b).

This reaction was verified after 12 hours with the techniques described. The dependence of the reaction on the pH level was low. In a pH range from pH 4 to pH 6, conversion rates totaling about 5% were measured for different isopulegol isomers after 20-hour incubation.

Here it was not critical whether the batches were incubated at RT, 30° C. or 37° C. The conversion was also not increased by addition of divalent ions, such as MgCl<sub>2</sub>, for example (cf. Test 1d). What was critical, however, was that the cell extracts, in the case of measurements above a pH of pH 5, either were dialyzed before the substrate was added, or EDTA was added to the batches, in order to suppress reduction of the citronellal substrate to citronellol by enzymes of the host. No effect of this treatment on the activity of the Zm-SHC-1 was found. Where this treatment was not carried out, the substrate was reduced almost completely to citronellol within 20 hours, and there was no longer any measurable cyclization to isopulegol. Zm-SHC-1 is therefore able to cyclize citronellal, but not citronellol, to isopulegol. It is very likely that unspecific dehydrogenases are responsible for the reduction reaction.

In order to rule out a chemical reaction being responsible for the cyclization, boiled-off cell extracts were used. In these controls and in controls with cell extracts from cultivation with empty vectors, however, no corresponding conversion was found (cf. Test 1a).

With (+/-)-citronellal as the substrate it was possible, following the reaction, to detect various isomers of isopulegol, which have not yet been precisely identified (cf. Tests 2 and 3). In order to verify whether these isomers originated from the different isomers of the starting substrate or if only one isomer was accepted as the substrate and was differently converted, the same studies were carried out with (+)-R-citronellal and (-)-S-citronellal. Here it was found that, depending on the substrate, different isopulegol isomers are formed. Interestingly, the conversion of (+)-R-citronellal took place from a pH of 4 to a pH of 7 without substantial differences, at a rate of about 5%. The enantiomer, in contrast, was converted with conversion rates of approximately 4.5% only up to a pH level of pH 6. Here as well, the conversion rate showed virtually no fluctuation in terms of the individual pH levels between pH 4 and pH 6.

Sequences:

SEQ ID NO: 1-326 nucleic acid/amino acid sequences of various SHC genes SEQ ID NO: 327-388 PCR primers

The disclosure of the publications cited herein is expressly referred to.

There follows a listing of SHC enzyme sequences which can be used in accordance with the invention:

Enzyme Sequences

```
>seq_ID 4
MNASRFSLKKILRSGSDTQGTNVNTLIQSGTSDIVRQKPAPQEPADLSALKAMGNSLHTLSS
ACEWLMKQKPDGHWVGSVGSNASMEAEWC LALWFLGLEDPHPLRPLGKALLEMQRDPDGS
WGTYYGAGSGDINATVESYAAALRSLGYAEDDPAVSKAAAWIISKGGLKNVRVFTRYWLALIGE
WPWEKTPNLPPEI IWFPDNFVFSIYNFAQWARATMMPLA ILSARRPSRLRPRQDRDLDFPGE
RANFDYELPTKEGRDVIADFFRLADKGLHWLQSSFLKRAPSREAAI KYVLEWI IWHQDADGGW
GGIQPPWVYGLMALHGEQYQPHHPVMAKALDALNDPGWRHDKGDASWI QATNSPVWDTML
SLMALHDANAERFTPEMDKALDWLLSRQVRVKGDWSVKLPNTEPGGWAFYANDRYDPTD
DTAVALIAIASCRNRPEWQAKGVBEAIGRGVRLVAMQSSCGGWGAFDKDNNKSLAKI PFCD
FGEALDPPSVDVTAHVLEAFGLGLPRDLPCIQRGLAYIRKEQDPTGPWFGRWGVNYLYGTGA
VLPALAALGEDMTQPYI SKACDWLINCQEQNGGWGES CASYMEVSS IGHGATTPSQTAWALM
GLIAANRPQDYEAIAKGCYRILDLQEEEDGSWNEEFTGTGFPYGVGQTI KLDDPAISKRLMQG
AELSRAFMLRYDLRQLFPI IALSASRLIKLGN
```

```
>seq_ID 2
MGIDRMNSLSRLMKKI FGAEKTSYKPSADTI IGTDTLKRPNRRPEPTAKVDKTI FKTMGNSLNN
TLVSA CDWLIGQQKPDGHWVGA VESNASMEAEWC LALWFLGLEDPHPLRPLGKALLEMQRD
DGSWGVYFGAGNGDINATVEYAAALRSLGY SADNPVLKKA AAWIAEKGG LKNIRVFTRYWLALI
GEWPWEKTPNLPPEI IWFPDNFVFSIYNFAQWARATMVPIA ILSARRPSRLRPRQDRDLDFPGE
GRARFDYELPKKEGIDLWSQFFRTTDRGLHWVQSNLLKRNSLREAAIRHVLEWI IRHQDADGG
WGGIQPPWVYGLMALHGEQYQPHHPVMAKALSALDDPGWRHDRGESSWI QATNSPVWDTM
LALMALKDKAEERFTPEMDKAADWLLARQVKVKGDWI KLDPVPEPGWAFYANDRYDPTD
DTAVALIALSSYRDKKEWQKKGVEDAITRGVNWLIAMQSECGGWGAFDKDNNRSILSKI PFCD
```

-continued

Enzyme Sequences

FGESIDPPSVDTVTAHVLEAFGLTGLSRDMPVIQKAIIDYVRSEQEAEAWFGRWGVNYIYGTGA  
VLPALAAI GEDMTQPYITKACDWLVAHQEQDGGWGESCSSYMEIDSIKGP TTPSQTAWALM  
GLIAANRPEDYEAIKAGCHYLIDRQEQDGSWKKEEFTGTGFPYGVGQTIKLLDDPALSKRLLQG  
AELSRAPMLRYDFYRQFPFIMALSRAERLIDLNN

>seq\_ID 5  
MTVTSSASARATRDPGNYQTALQSTVRAAADWLIANQKPDGHVWGRAESNACMEAQWCLAL  
WFMGLEDHPLRKRLLQSLDSQRPDGAWQVYFGAPNGDINATVEAYAALRSLGFRDDEPAVR  
RAREWI EAKGGLRNIRVFTRYWALIGEWPEWKT PNI PPEVIWFPPLWFFPSIYNFAQWARATLM  
PIAVLSARRPSRPLPENRLDALFPHGKAFDYELPVKAGAGGWD RFRGADKVLHKLQNLGN  
RLNLGLFRPAATSRVLEWMIHQDFDGAWGGIQPPWIYGLMALYAEGYPLNHPVLAKGLDALN  
DPGWRVDVGDATYIQATNSPVWDTI LTLA LAFDDAGVLGDYPAEAVDKAVDWWLQQRVVRPGDW  
SMKLPHVKPGGWAFYANNYPDTDDTAVALI LALAPLRHDPKWKAKGIDEA IQLGVDWLI GMQ  
SQGGWGA FDKDNNQKI LTKI PFC DYGEALDPPSV DVT AHI IEAFGKLGISRNHPSMVQALDYI  
RREQE P SGPWFGRWGVNYVYGTGAVLPALAAI GEDMTQPYI GRACDWLVAHQEQDGGWGE  
SCAS YMDI SAVGRGTTTASQTAWALMALLAANRPQDKDAI ERGCMWLVERQSAGTWDEPEF  
TGTGFPYGVGQTI KLNDPALSQLMQPELSRAPMLRYGMYRHYFPLMALGRALRPQSHS

>seq\_ID 78  
MTLTSASARAPRDPGNYQTALQSTVRAAADWLIANQKPDGHVWGRAESNACMEAQWCLAL  
WFMGLEDHPLRKRLLQSLDSTQRPDGAWQVYFNAPNGDINATVEAYAALRSLGYPDSEPAVR  
RAREWI EAKGGLRNIRVFTRYWALIGEWPEWKT PNI PPEVIWFPPLWFFPSIYNFAQWARATLM  
PIALLSARRPSRPLPENRLDALFPFRGRDADYELPVKANAGGWDKFRGADKVLHQLALQFN  
RLNLGLFRPAATSRVLEWMIHQDFDGAWGGIQPPWIYGLMALYAEGYPLNHPVLAKGLDALN  
DPGWRVDVGEATYIQATNSPVWDTI LTLA LAFDDAGVLGDYPAEAVDKAVNWWLQQRVVRPGDW  
SMKLPHVKPGGWAFYANNYPDTDDTAVALI LALAPLRHDPKWKAKGIDEA IQLGVDWLI GMQ  
SQGGWGA FDKDNNQKI LTKI PFC DYGEALDPPSV DVT AHI IEAFGKLGISRNHPSMVQALDYI  
RKEQEP SGPWFGRWGVNYVYGTGAVLPALAAI GEDMTQPYI GRACDWLVAHQEQDGGWGE  
SCAS YMDI SAVGRGTTTASQTAWALMALLAANRPQDKDAI ERGCMWLVERQSAGTWDEPEFT  
TGTGFPYGVGQTI KLNDPALSQLMQPELSRAPMLRYGMYRHYFPLMALGRALRPQSHS

>seq\_ID 209  
MDSILAPRADAPRNIDGALRESVQQAADWLVANQKPDGHVWGRAETNATMEAQWCLALWFL  
GLEDHPLRVRVLRGALLDQRPDGAHVYFGAPNGDINATVEAYAALRSLGHRDDEEPLRKR  
DWILSKGGLANIRVFTRYWALIGEWPEWKT PNI LPEVIWLPWFPSIYNFAQWARATLMP IAV  
LSAHRPSRPLAPQDRDLALFPQGRDSFNVDLPARLGAGVWDVIFRKIDTILHRLQDWGARRGP  
HGIMRRGAI D HVLQW IIRHQDYDGSWGGIQPPWIYGLMALHTEGYAMTHPVMAKALDALNBP  
WRIDIGDATFIQATNSPVWDTMLSLA LAFDDAGLGERYPEQVERAVRVLKRVQLVPGDWVSKL  
PDVKPGGWAFYANNYPDTDDTSVALMALAPFRHDPKWKQAEIGDAIQRGIDWLVAMQCKE  
GGWGA FDKDNDKKI LAKI PFCDFGEALDPPSADVTAHI IEAFKAVGLDRNHPSIVRALDYLKRE  
EPEGPWFGRWGVNYVYGTGAVLPALAAI GEDMRQPYIARACDWLIARQQANGGWGESCVSY  
MDAKQAGEGTATASQTAWALMALIAADRPQDRDAIERGCLYLETQDRDGTQEVHYTGTGFP  
GYGVGQTI KLNDPLLSKRLMQPELSRFMLRYDLYRHYFPMMAIGRVLQRGDRSGH

>seq\_ID 193  
MNVIRQLNSGVNAKSLDDGIESAIEWLAENQDKEGFWGMLESNSCIEAEWILAMHLLGVKD  
DPKYDKVVQAILNEQREGDSWAVYDAPAGDINATVEAYAALRTAGFGAGDERLIKARNWIFS  
HGGLKNRVVFTRYWALIGEWPEWDETPALAPEI IYLPACPLNIYDFACWARATLVPLSVLSVR  
RPVKPLPAESRLDELPEGRENADYSLPESKGLAERFFLVVDWFLKKNRLLPMQFGREKAI R  
LCLLEWIVRHQDYDGGWGGIQPPLIYSLIALNTEGYINHPVLSKGLDAFNPPWAYEKNGGVLQ  
CSES PVWDTLFTMLALFESGCSFDDTPMMPALDWILSKQITSWGDWQVKVRGVRPGGWAF  
ERANTAYPDVDDTALALVVLAEARRHVKDSAAVDAALERA EEWILGQLCRNGGWAAPDRDNN  
SAIVTKI PFCDFGEVLDPPSV DVT AHVVEALAALGRDRHDPVVARALKYIRSEQEPGGSWFGR  
WGVNHI YGTCAVLPALAAI GEDMRAPYVLRADWLVRHQNDGGWGESAS YMDSSQCGQ  
GSSTASQTGWALMALVAMS SHDYDEAIRRGLDYLSSHQSGTWDEPQY TGTGFPYGVGGER  
TNLKEAGATLDQGCBLARGFMINYNMRYHYPFLI AMARARRHLGLAANPRHQDSRS SVEVAPE  
ALRGRACC

>seq\_ID 246  
MRRLDTFPPEIPTGSRDKPPSGEEHSCSTPAEPLRSRLDEGILRAVDWLVCDQHPDGFWAGM  
LQNSCMCAEWVLA MHFLGIDDDPKYDGVIRAILGEQRADGSGWGFHKA PNGDINTTVECYAA  
LRASGLAPESAPLSARAWILAGGGLANIRNFTKYWALIGEWPEWGTPTI PPELIFFPFRPLM  
IYHFASWARSTI VPLSILSARRPVRPLPEDRRLDELFPQGRSAFDRLPKDKGWL SWEGFPFHC  
DRILRLYARTRRAPRETAIRVCLLEWIRREQTDGAWSGIQPPWIYALLALHAEGYGLDHPILRA  
GLRFPD SHWSYERGGIYLAQASESPVWDTVLSLRLADCGEERKASVSIASALEWLLNRQISV  
PGDWAVRVPSPVPCGGWAFGRANSFYPDVDDTAVAI EVLARLRPPTANQS AVDRAIRSARDWV  
LAMQCSNNGGWA AFRDNDFKLVTKI PFCDFGELLDPSPVDVTAHVIEALAA LGWDMT SREIEA  
AVSFIRREQEAGSWFGRWGVNHI YGTATVLPALRAI GEDMSSAYVLRADWLASRQNDAGG  
WGETPAS YMDSLR VGVGESTASQTAWALMGLVAVGSGAHDDTVRRGIDFLLFAHQGGTWE  
PQY TGTGFPYGVGERI RLDMGASLQGTLEQRAFMINYNLRYHYPFLMALGRARYHLQLRR  
SAREGGNETTPNGSAL

>seq\_ID 151  
MKISKNPI SHALTSFNDAARETADNSAARKSGKIHHLPATIWKKKESTVSSPLDIAIERTQEFFFR  
EQLPAGYWWAELESNATI TAEYIMLFHFMGLVNREKERKMANYLRLRQQTTEGYWTIWHGGPG  
DLSTTI EAYFALKLAGYPADHPSMSKARAFIEHGGILKARVFTKIPLALPGEFVSWLGVPSMPIEM  
MLLPAGFTFNMYPSSWSRATI I PLSI VMAERPVRKLP PWARVQELVYRPPRPTDYTF TKEDGIL

## Enzyme Sequences

TWKNIFIGIDHVLKVYEASPIRPGRKKAMAI AEKWVLEHQEPTGDWGGIQPAMLNSVLALHVLG  
YANDHPAVAKGLQALANFCIEGEDELVLQSCVSPVWDTALGLMAMVDSGVPTDHP SLSKAAQ  
WLLDREVRRPGDWKIKCPDLEPGGWAFFEMNDWYPDVDDSGIVMMAIKNVKVKDQRAKEDTI  
TRGI AWCLGMQSKNGGWGAFDKDNTKHLNKI PFADLEALIDPPTADLTGRMLELMGT YGYPK  
DHAAVRAALQFVKEQEPDGPWWGRWGVNYIYGTWSVMSGLA AFGEDMSQPWIRKAVDVLV  
EHQNEDEGGWGECCESYADPRLAGVGPSTASQTGWALLTLAAGEVASSSVVRGVQYLLDTQ  
KPDGTANDEDAFTGTGFPKPFMIKYHIYRNCFPPLMALGRYRTLAKGKL

>seq\_ID 142

MKSRKYPI SHALTSFNHTTVAPEAPAPISVKSPAKVHRLPSSIWKKMEGSAGNPLDKAVELTR  
DFFFRQLPDGYWVAELESNTITAEYIMLFHFLGMVDKDKERKMANVLLRQQTEEGYWTVM  
HNGPGDLS TTIEAYFALKLAGYHADHIALRKARDFILANGGILKSRVFTKTF LAMFGEFSWLGVP  
SMPIELMLLPDWAYLNVYEFSSWARATIIPMSVLMANRPVYKLP PHARVQELYVRPPRPTDYTF  
TKEDGIFSLKNFFIGVDHLLKIIYESSPIRPFKRATEKVEQWILEHQEKTGDWGGIQPAMLNAI LA  
LHCLGYANDHPAVAKGLEALANFTI ESDSLVLQSCISPVWDTALVLQAMQEA SVPLDHP SLIK  
ASQWLLDREVRIKGDWIKSPDLEPGGWAFFQNDWYPDVDDSTAVMIAIKDIKVKNTKARQD  
AIRRGIDWCLGMQSKNGGWGAFDKDNTKHLNKI PFADLEALIDPPTADLTGRMLELMGNFY  
TKDHPQAVSALEFLKNEQEPGPWFGRWGVNYIYGTWVVLIGLEAIGEDMNSPYIKKSVNWI K  
SRQNLDDGGWGEVCDSDYDRITLMGCGPSTASQTSWALMALMAAGEVGCQAVERGIQYLLAT  
QNSDGTWDEEAFGTGFPKPFMIKYHIYRNCFPPLTALGRYRRLTAGTHAQ

>seq\_ID 152

MNSCKHPI SHALTSFNGETADA AKKQPVKPGAKI HHLPASIWKKKEGESKSP LDIAIENSRDFF  
REQLPDGYWVAELESNTITAEYIMLYHFMGIVDQERERKMATYLLSKQTAE GFWTIYFGGPG  
DLSTTVEAYFALKLAGYPADHPAMAKARAFILDNGGIIKCRVFTKIFLALFGEFAWFGVPSMPIEL  
ILLPNWAYFNMYELSSWSRATIIPLSIVMTERPVRKLPSPSRVQELYVRPPRIDYTF SKEDGIIT  
WKNFFIGVDHILKVIYESNPIRPFKRALATAENWVLDHQESTGDWGGIQPAMLNSVLALHCLG  
YANDHPAVAKGLEALANFCIETEDSLVLQSCISPIWDTALALKALVDSVPTDHPALVKAQWLL  
DKEVRRPGDWKIKCPLESGGWAFFELNDWYPDVDDSGFVMMALKDVAVKDRKSM DGAIKR  
GINWCLGMQSKNGGWGAFDKDNTKYL NKI PFADLEALIDPPTADLTGRMLELMGT FGYSKDY  
PAAVRALEFIKKNQEPGSWWGRWGVNYIYGTWSVLGGLAAGIEDLNQPIRKAVNWLKSRQ  
NMDGGWGETCESYHDTSLAGIGESTPSQTGWALLSLMSAGEANSS TVARGIQYLIANQKSDG  
TWDEEQYTGTFPKPFMIKYHIYRNCFPPLTALGTYRKL TGGMA

>seq\_ID 146

MTSPFKHPISNALTSFNNGNFAEPEQCVEQQTGAKVHHLPASIWKRKMGKAKSPLDVAIEGSRD  
FFQQLPKGYWVAELESNTITAEYIMLFHFLGLVDRERQRKMSNYLLSKQTEEGFWPIY YG  
GPGDLS TTIEAYFALKLAGYPADHPALAKARAFILEQGGVVKSRVFTKIFLALFGEFEWQGVPS  
MPVELNLLPDWAYINIYEFSSWARATIIPLSVVMHSRPPVRRVPPSARVQELFVRQPTAADYSFA  
KNDGIFTWENFFLGLDRVLKVYKESPLRPFKNMALAKAEWVLEHQEPTGDWGGIQPAMLNA  
VLALNVLYGQNDHPAVEGQLRALANFCIETEDQLVLQSCVSPVWDTALALKALLDAGVPPDHP  
SLVKAQWLLDKEVTRPGDWRVKS PALEPGGWAFFELNDWYPDVDDSGFVMIALKGIQVKDR  
KSM DAAIKRGINWCLGMQSKNGGWGAFDKDNTRHV LNKI PFADLEALIDPPTADLTGRMLELM  
GTFNYPITLPAQR AIEFLKKNQEPGPWWGRWGVNYIYGTWSVLCGLAAGIEDMDQPIYRKA  
VNWI KSRQNLDDGGWGETCQSYHRTLAGVGES TPSTGWALLGLLAAGEMHSATVVRGVQY  
LISTQNSDGTWDEQQTGTGFPKPFMIKYHIYRNCFPPLMALGTYRRTLRTQTQ

>seq\_ID 147

MSPCKHPI SHALTSFNGETADSVPVQTPKTGAKI HHLPPSIWKKKEGELKSP LDIAIENSRDFF  
REQLPDGYWVAELESNTITAEYIMLYHFMGLVDRERERKMANVLLSKQTEEGFWTIY YGGP  
GDLS TTIEAYFALKLAGYPADHPAMVKARAFILDNGGIIKTRVFTKIFLALFGEFAWFGVPSMPIE  
LILLPNWAYFNMYELSSWSRATIIPLSIVMTRPVRKLPSPSRVQELYVRPPSPIDYTF TKEDGIF  
TWKNFFIGVDHILKVIYESNPIRPFKAMLAENWVLEHQEATGDWGGIQPAMLNSVLALHCL  
GYANNHPAVAKGLEALANFCI ESDSLVLQSCISPVWDTALALKALVDSV PNDHPALVKAQ  
WLLDKEIRKAGDWKIKAPNLEPGGWAFFELNDWYPDVDDSGFVMMALKDVAVKDRKSM DTAI  
KRGISWCLGMQSKNGGWGAFDKDNTKYL NKI PFADLEALIDPPTV DLTGRMELMGT FGYAK  
DYPPAVRALDFIKRNQEPDGSWWGRWGVNYIYGTWSVLCGLSAGMEDLNQPIYRKA INWLK  
RQNLDDGGWGETCESYHDTSLAGI GASTASQTGWALLALMAVGEENASAVARGVQYLLATQKS  
DGTWDEDLYTGTGFPKPFMIKYHIYRNCFPPLTALGTYRRTGGR AEMQVSEHNK

>seq\_ID 144

MKISKHPI SHALTSFNETAKEKTEEPQKRGKVVHHLPASIWKKRDVETTSPLDQAIKRSQEFFL  
REQLPAGYWVAELESNTITAEYVILFHFMLVNRDKDRKMATYLLSKQTEEGCWC IWHGGP  
GDLS TTIEAYFALKLAGYPADHPAMQKARTFILGKGGILKARVFTKIFLALFGEFSWLGVP SMP IE  
MMLLPNGFTFNLYEFSSWSRATIIPLSIVMAERPVRKLPWARVQELYVRPPRPMDYTF TKEDG  
ILLTWKNIFIGIDHILKVYEASPIRPGMKAMAI AEQVLDHQEPTGDWGGIQPAMLNSVLALHCL  
GYANDHPAVAKGLQALANFCI ESDDEI VLQSCISPVWDTALALMAMVDSVPTDHPALVKAQ  
WLLDREVRRKVDWKIKAPNLEPGGWAFFQNDWYPDVDDSGIVMMAIKDVKVKDSKAKAEAI  
QRGI AWCLGMQSKNGGWGAFDKDNTKHLNKI PFADLEALIDPPTADLTGRMLELMGT FGYPK  
DHAAVRAALQFVKEQEPDGPWWGRWGVNYIYGTWSVLCGLKAYGEDMGQP YVRKAVEWL  
AAHQNPDDGGWGECCESYCDQKLAGTGPSTASQTGWALLSMLAAGD VDHPAVARGIRYLIETQ  
QPDGTWDEEDQFTGTGFPKPFMIKYHIYRNCFPPLMAMGRYRALKGHKG

>seq\_ID 15

MAEQLV EAPAYARTLDRAVEYLLSCQKDEGYWGP LLSNVMTAEAYVLLCHILDRVDRDRME  
KIRRYLLEHQREDDGTWALYPGGPPDLDTTIEAYVALKYIGMSRDEEPMQKALRFIQSQGGI ESS  
RVFTRM WLLALVGEY PWEKVPMPVPEIMPLGKRMLNIYEFGSWARATVVAISIVMSRQVFPFL

-continued

Enzyme Sequences

PERARVPELYETDTPVPPRRRGAKEGGGRIFDALDRALHGYQKLSVHPFRRAAEIRALDWLLERQ
AGDGSWGGIQQPPWFYTLIALKILDMTQHPAFIKGWEGLLEYGVLDLYGGWFMFQASISPVWDT
GLAVLALRAAGLPADHDLRVKAGEWLLDRQITVPGDWAVKRPNLKPGGFAPQFDNVYYPDVT
DTAVVVWALNSLRLPDERRRRDMTKGFRWIVGMQSSNGGWGAYVDVNTSDLPNHIPFCDF
GEVTDPPSEDEVTAHVLECFGSFGYDDAWKVIIRRAVEYLRKREQKPDGGSWFGRWGVNLYLGT
AVVPALKAVGIDVREFFIQKALDWVEQHQNPDDGGWGEDCRSYEDPAYAGKGASTPSQTAWA
LMALIAGGRAESDSVRRGVQYLVEVETQRPDGGWDEPYTGTGFPDFYLYGTYMRHVFPPTLAL
GRYKQAIERR

>seq\_ID 16
MAEQLVPEAPAYARTLDRAVEYLLSCQKDEGYWVGPLLSNVTMEAEYVLLCHILDRVDRDRME
KIRRYLLHEQREDGTWALYPPGGPPDLDTTIEAYVALKYIGMSRDEEPMQKALRFIQSQGGIESS
RVFTRMMLALVGEYVPEKVPMPVPEIMPLGKRMPLNIEYFGSWARATVVALSIVMSRQVPFPL
PERARVPELYETDTPVPPRRRGAKEGGGWIIFDALDRALHGYQKLSVHPFRRAAEIRALDWLLER
QAGDGSWGGIQQPPWFYALIALKILDMTQHPAFIKGWEGLLEYGVLDLYGGWFMFQASISPVWD
TGLAVLALRAAGLPADHDLRVKAGEWLLDRQITVPGDWAVKRPNLKPGGFAPQFDNVYYPDVT
DDTAVVVWALNTRLRPLPDERRRRDMTKGFRWIVGMQSSNGGWGAYVDVNTSDLPNHIPFCD
FGEVTDPPSEDEVTAHVLECFGSFGYDDAWKVIIRRAVEYLRKREQKPDGGSWFGRWGVNLYLGT
GAVVSALKAVGIDTREPYIQKALDWVEQHQNPDDGGWGEDCRSYEDPAYAGKGASTPSQTAW
ALMALIAGGRAESEARRGVQYLVEVETQRPDGGWDEPYTGTGFPDFYLYGTYMRHVFPPTLAL
LGRYKQAIERR

>seq\_ID 141
MTSPFKHPISNALT SFNGNVAEPEQSV EQSGAKVHHL PASIWKRMGRAKSPLDVAIEGSRD
FFQEQPLPKGYWVAELESNTIATAEYIMLPHFLGLVDPERQRKMS TYLLSKQTEEGFWTIYYG
GPGDLSTTIEAYFALKLGSYEPEDHPALAKARAFILEQGGVVKSRVFTKIFLALFGEFDWQGI PSM
PVELNLPLDWAYINIYEFSWARATIVPLSVMHSRPPVRRVPPSARVQELFVRQPTAADYSFAK
NDGLFTWKEKFLGLDRVLKVEKSPLRPFKKTALAKAEWVLEHQEPETDGGGIIQAPMLNAIL
ALNVLGYNRNDHPAVEQGLRALANFCIETEDQLVLQSCVSPVWDTALALKALLDAGVPPDHPSL
VKGAQWLLDKEVTRAGDWRVKS PNLEAGGWAFELNDWYVDVDDSGFVMIALKGIQVKDKH
AMDAAIKRGINWCLMGQSKNGGWGAFDKDNTKXVLNKI PFADLEALIDPPTADLTGRMLELMG
TFDYVPTFPAAQRAIEFLKKNQEPPEGPWWRWGVNLYLGTWSVLCGLAAIGEDMDQPYIRKA
VNWIKSRQNDGGWGETCQSYHRTLAGVGESTPSQTGWALLSLLAAGEMHSATVVRGVQYL
ISTQNSDGTWDEQQTGTGFPKPYMIKYHIYRNCFPPLMALGTYRTLTRTQP

>seq\_ID 195
MNPACYKISSLSLNAEPVEQAPLPAKRTGSKVHRLPSSIWKVMVAEAKSPLDKGIERTRDFP
LREQLPDGYWVAELESNTIATAEYVMLPHFLGMVDRERERKLANYILAKQTEGFWSLWHNG
PGDLSTTIEAYFALKLAGYSADHPAMAKARAPVLANGGIIKARVFTKIFLALFGEFAWFGVPSMP
ELMLLPDWAYFNMYEFSWSRATIIPLSVMMSERPVKLPRAAQVQELFVRPPTDYITITRED
GLFTWKNFFI GADHLIKVYESPIRPFKRAVALAENWILEHQEQSGDWGGI QPAMLSILALHC
LGYANDHPAVAKGLDALANFCIEDDDCIVLQSCVSPVWDTALALVALQEQADVPADHPALVKAA
QWLLNLLEVRRKGDWQVCKPELEPGGWAFELNDWYVDVDDSGFVMLSINKI KVRDRKHREE
AIKRGIAWCLMGQSENGGWGAFDRNNTKYLNLKI PFADLEALIDPPTADLTGRMLELMGNFDY
PKSHPAABRALAFLKKEQSESGPWWRWGVNLYLGTWSVLCGLEAIGEDMNQPYIRKAVNWI
KSRQNDGGWGEVCESYFDRSLMGGPSTASQTGWALLALMAAGANSRAAAQGVKYLLET
QNEGTWDEDAFTGTGFPKPFMIKYHIYRNCFPPLTALGRYRRLTAAG

>seq\_ID 3
MTATTDGSGTASLRPLAASDTDITIPAAAAGVPEAAARATRRATDFLLAKQDAEGWKGDL
ETNVMTDAEDLLLRQFLGIDQEEETRAAALFIRGEQREDGTWATFYGGPGLSTTIEAYVALRL
AGDSPEAPHMARAEEI RSRGGI ASARVFTRIWLALFGWKKWDDLPELPELIYFPTWVPLNI
YDFGWARQTI VPLTIIVSAKRFPVRPAPFPLDELHDTDPARPNNPRLPAPVASWDGAFQRIKALH
AYRKVAPRRLRRAAMNSAARWIIERQENDGCWGGIQQPPAVYSVIALYLLGYDLEHPVMRAGLE
SLDRFAVWREDEGAMI EACQSPVWDTCLATIALADAGVPEDHPQLVKASDWMLGEQIVRPGD
WSVKRPLPGLPPGGWAFEFHNDNYPDIDDTAEVVLALRRVRRHHPERVEKAI GRGVRWNLGMQ
SKNGAWGAFVDVNTSAPFNRLPFCD FGEVIDPPSADVT AHVVEMLAVEGLAHDPRTRRGIQW
LLDAQETDGSWFGRWGVNLYLGTGVSIPALTAAGLPTSHPAIRRAVRWLESVQNEDEGGWGE
DLRSYRYVREWSGRGASTASQTGWALMALLAAGERD SKAVERGVAVLAAATQREDGSWDEP
YFTGTGFPWDFSINYNLYRQVFPPLTALGRYVHGEFFAKKPRADAPAEAPAEVKG

>seq\_ID 18
MTKQLLDTMPVQATLEAGVAHLLRRQAPDGYWVAPLLSNVCMEEAYVLLCHCLGKKNPEREA
QIRKYIISQRREDGTWYIPGGPSDLNATVEAYVALKYLGEFASDPQMVQAKEFIQNEGGIEST
RVFTRLWLAMVGGYQVWDLKLPVIPPIMHLPKSVPLNIIYDFASWARATIVTLSYRHESPTCDTAS
GLCKGSGIVRGEPPKRRSAKGGDSGFVALDKPLKAYNKWPIQPRKSGEQKALEWILAHQ
EADGCWGGIQQPPWFYALALKCLNMTDHPAFVKGFEGLEAYGVHTSDGGWFMFQASISPIWDT
GLTVLALRSAGLPPDPHPALIKAGEWLVSKQILKDGWVRRRKA KPPGGWAFEFHCENYPDVD
DTAMVVLALNGIQLDEGKRRDALTRGFRWLRREM QS SNGGWGAYVDVNTRQLTKSDSIFATS
GEVIDPPSEDEVTAHVLECFGSFGYDEAWKVIIRKAVEYLRKREQKPDGGSWFGRWGVNLYLGTG
VVPGLKAVGVDMREPWQKSLDWLVEHQNEDEGGWGEDCRSYDDPRLAQGVSTPSQTAW
ALMALIAGGRVESDAVLRGVTYLHDTQRADGGWDEEVYGTGFPDFYLYGTYMRHVFPPTLAL
LGRYQEQMQRIRG

>seq\_ID 245
MNPIRGKRGSAADFLEEEYQWENLADHGESGRTPGGHPAALKEYEAGSATEHTGHHCVHH
LGVRNSWLRKI EKAIDNACQGLFKTQYEDGYWWS ELESNTITSEYIMLLYLLVSRPEQQKSM

-continued

## Enzyme Sequences

VKYLNLQQRPDGWSGLYYDGGNLS T T I E A Y F A L K L A G E H C E S E P M R R A R E F I L S K G G I E S A R  
 V F T K I W L A L F S Q Y D W D K V P S M P V E L V L L P S S L Y F N I Y E F S S W A R G T V V P L S I V M S I R P R C P L P A K  
 C S I K E L Y V P G S K H K N F A S C T H K L F P L F D R I A K A P E R R P V P S L R N K A V Q A A E T W V L D H Q E D S G D  
 W G G I Q P M V Y S V L A L Y L Y G P L D H E V I V K G I K A L D A F C M E D E E G T R M Q S C V S P V W D T A L T V L S  
 M L D A G V A A E H P G L E K A G R W L L E N Q V L T G G D W Q I K N D S L P G G W A F E F Y N T R Y P D V D S A V V L  
 S T L N R F N A E R V E G L E F A K C R G M E W C L S M Q S S N G G W A A F D K D N T L E I L N R I P F A D Q E A M V D Y P  
 T A D V T G R V L E A M G Y L G Y D G S H P R A R K A I Q F L K K R Q E R D G C W W G R W G V N Y I Y G T W S V L K G L I  
 S I G E D P R A A Y T R A A V R W V K D H Q N S D G G W E T C E S Y E N P E L R G Q G P S T P S Q T A W A L M S L I A C G  
 E M K S Q E A S R G I Q Y L L R T Q K R D G T W E E L H F T G T G P K H F Y I R Y H N Y R N C F P L M A L G Q Y L R A L E R

&gt;seq\_ID 221

M T A T T D G S T G A L P P R A A S A S E P H D T I P Q A A G S V G I Q D A A A R A T Q R A T D F L L S R Q D A E G W W K G  
 D L E T N V T M D A E D L L R Q F L G I Q D E K T T R A A G L F I R G E Q R A D G T W A T F Y G G P G D L S A T I E A Y V A L  
 R L A G D G P D E P H M A K A S A W I R E R G G I A S A R V F T R I W L A L F G W W K W D D L P E L P P E L I Y F P K W M P  
 L N I Y D F G C W A R Q T I V P L T V V S A K R P V R P A P F P L D E L H A D A N D P N P A K P L A P M V S W D G L F Q R L D  
 V A L H T Y R K V A P R R L R K A M N T A A R W I I E R Q E N D G C W G G I Q P P A V Y S V I A L Y L L G Y D L E H P V M R  
 E G L A S L D R F A V W R D D G A R M I E A C Q S P V W D T C L A T I A L A D A G V P A D H P Q L V R A A D W M L G E E I V  
 R P G W A V K R P Q L P P G G W A F E F H N D N Y P D I D D T A E V V L A L R R V K H H D P E R L D N A I R R G V R W N L  
 G M Q S K D G G W A F D V D N T S P F P N R L P F C D F G E V I D P P S A D V T A H V V E M L A F E G L S H D P R T R R  
 G I Q W L L S A Q E A N G S W F G R W G V N Y Y G T G S V V P A L V A A G L P A S H P A I R R A V T W L E T V Q N D D G  
 G W G E D L R S Y P E A A E W S G K G A S T A S Q T G W A L L A L L A A G E R S K A V E R G I E W L A Q T Q R P D G S W  
 D E P Y F T G T G F P W D F S I N Y H L Y R Q V F P L T A L G R Y V N G E P L V E V K G G

&gt;seq\_ID 160

M K G K E P T R E E L L S F S S G I Q M D S S A E N T T P V S T E E L Q E K V R L A A E S L I S R Q E E G Y W V E P L E A D  
 V T I T S E Y I L L Q Y L L G R E R D E F F R R A A P F I L E S Q G E D G G W P L Y H G G P A E I S A T V K A Y L A L K L L G Y D  
 A D H P A M Q R A R A L V L E R G G A I N V N V F T R I T L A L F G Q Y D W K G V P A L P P E M I L L P R W F P L S I Y T V S Y  
 W S R T V I V P L L F I Y H Y K P L L E L P P E K G V Q E L F I T P M S E V R V H Y A W D K H W V S W K N L F F V L D R I L Q A  
 W N R H P P S F L R R K A L K A M E W M I P R L K G E G L G A I Y P A M A N S V L A L R L E G Y A M D H P L V R R A I Q A  
 I D D L V F D L G E Q Q S V Q P C H S P I W D T A L A L G A L Y E A G L D E G S P F V S R A L D W F C R K E V R T V G D W S  
 V R V P G V E A G G W A F Q F E N D Y P D I D D T S V V L M D F A K W V P E M G A Y R D V F R R A I E W T L S M Q G T D  
 G G W A F D K D N D F L F L N N I P F A D H G A L L D P S T S D V T G R V T E L L G I L G Y D A R T P V V R R A L R F L R K E  
 Q E E N G S W Y G R W G V N Y I Y G T W S V V S A L K A V G E D M S A P Y V Q K A M Q L F S R Q N P D G G W G E S C Y  
 S Y F R K D T A G E G V S T S S Q T A W A L I A L I H G G H V R H P A V S K G I D F L L S R Q Q A D G K W L E Q E Y T G T G F  
 P K V F Y L R Y N M Y R D Y F S L W A L S L Y R N V L L D G Q S R V E R L A R R W K G N P Y P V R S R F L A

&gt;seq\_ID 161

M E G K D P T R E E L L S F T S G I Q M D S R V G N T N P V S T E E L Q E K V R L A A E S L I S R Q E E G Y W V E P L E A D  
 I T T S E Y V L L Q Y L L G R E R D E F F R R A A P F I L E S Q G E D G G W P L Y N G G P A E I S A T V K A Y L A L K L L G Y D  
 A D H P A M Q R A R A L V L E R G G A I N V N V F T R I T L A L F G Q Y D W K G V P A L P P E M I L L P R W F P L S I Y T V S Y  
 W S R T V I V P L L F I Y H Y K P L L E L P P E K G V Q E L F I T P M S E V R V H Y A W D K H W V S W K N L F F V L D R I L Q A  
 W N R H P P S F L R R K A L K A M E W M I P R L K G E G L G A I Y P A M A N S V L A L R L E G Y E M D H P L V R R A I Q A  
 I D D L V F D L G E Q Q S V Q P C H S P I W D T A L A L G A L Y E A G L D E G S P F V S R A L D W F C R K E V R T V G D W S  
 V R V P G V E A G G W A F Q F E N D Y P D I D D T S V V L M D F A K W V P E M G A Y R D V F R R A I E W T L S M Q G T D  
 G G W A F D K D N D F L F L N N I P F A D H G A L L D P S T S D V T G R V T E L L G I L G Y D A R T P V V R R A L R F L R K E  
 Q E E N G S W Y G R W G V N Y I Y G T W S V V S A L K A V G E D M S A P Y V Q R A M Q L F S R Q N P D G G W G E S C Y  
 S Y F R K D T A G E G V S T A S Q T A W A L I A L I H G G H V R H P A V S K G I D F L L S R Q Q A D G K W L E Q E Y T G T G F  
 P K V F Y L R Y N M Y R D Y F S L W A L S L Y R N V L L D G Q S R V E R L S R R W K G T P Y P V R S R F L A

&gt;seq\_ID 240

M H E G E A M T A T T D G S T G A L P P R A A A A S E T H L D T P V A A G I Q E A A V R A V Q R A T E H L L A R Q D A E G W  
 W K G D L E T N V T M D A E D L L R Q F L G I Q D A A T V E A S A R F I R G Q Q R D D G T W A T F Y G G P G E L S T T I E A  
 Y V A L R L D G D A P D A P H M A K A S A W I R A Q G G I A A A R V F T R I W L A L F G W W K W D D L P E L P P E L I Y F P K  
 W A P L N I Y D F G C W A R Q T I V P L T I V S A K R P V R P A P F P L D E L H A D P A D P N P A K P L A P V A S W D G A F Q  
 R L D K A M H Q L R K V A P R R L R R A M N S A A R W I I E R Q E N D G C W G G I Q P P A V Y S V I A L H L L G Y D L Q H P  
 V M R A G L E S L D R F A I W R E D G S R M I E A C Q S P V W D T C L A T I A L V D A G V P A D H P Q L V K A A D W M L G E  
 E I V R P G D W S V K R P Q L P P G G W A F E F H N D N Y P D I D D T A E V V L A L R R V R H H D P D R V E N A I G R G V R  
 W N L G M Q S K N G A W A F D V D N T S P F P N R L P F C D F G E V I D P P S A D V T A H V V E M L A V E G L S H D P R T  
 R R G I E W L L A E Q E P D G S W F G R W G V N Y I Y G T G S V V P A L T A A G L P A S H P A I R R A V A W L E K V Q N D D  
 G G W G E D L R S Y K V K E W S G R G A S T A S Q T A W A L M A L L A A G E R D S K A V E R G V E W L A S T Q R A D G  
 S W D E P Y F T G T G F P W D F S I N Y H L Y R Q V F P L T A L G R Y V H G E P P S R T E A L

&gt;seq\_ID 231

M T A T T D G S S G P V R A G A A T A G D T T T T A A R T T A P G T D V R E A A G R A A E R A V E H L L A R Q D A Q G W  
 W K G D L E T N V T M D A E D L L R Q F L G I Q D A A T V E A S A R F I R G Q Q R D D G T W A T F Y G G P G E L S T T I E A  
 Y V A L R L A G D R P D D P H M Q R A A S W V R S R G G I A A A R V F T R I W L A L F G W W K W D D L P E L P P E L I L L P K  
 W V P L N I Y D F G C W A R Q T I V P L T V V S A K R P V R P A P F A L D E L H T D P A M P N P Q K R F A P A A S W D G F F  
 Q R A D K A L H L Y H K V A P R R L R R A M N A A A R W I I E R Q E N D G C W G G I Q P P A V Y S V I A L H L L G Y D L E H  
 P V M R A G L E S L D R F A V H R E E E G L P V R M I E A C Q S P V W D T C L A T I A L A D A G L P A D H P A L V K A A D W M  
 L S E Q I V R P G D W A V R R P G L P G G W A F E F H N D N Y P D I D D T A E V I L A L R R V K H P D P E R V E A A V A R G  
 T R W N L G M Q S L N G A W A F A D N T S P F P N R L P F C D F G E V I D P P S A D V T A H V V E M L A H E G M A E D  
 P R T R R G V R W L L R E Q E A N G A W F G R W G V N Y Y G T G A V V P A L I A A G L P A S H P S V R R A V T W L E S V  
 Q N E D G G W G E D L R S Y R E E Q S I G R G A S T A S Q T G W A L L A L L A S A G E R D G R A V E R G V A W L A R T Q R P  
 D G S W D E P Y F T G T G F P W D F S I N Y H L Y R Q V F P L T A L G R F L H G E K P V G R A A A R E G G

## Enzyme Sequences

&gt;seq\_ID 227

MTATTDGSGAANPSEATAHDPDTTTAADDLVAARRAAERSVEHLLGRQDEQGWKGD  
 ATNVTMDAEDLLLRQFLSIQDPETTRAAALFIRGEQLGDGTWNTFYGGPGDLSATI EAYVALRL  
 AGDRPDEPHMARAAGNIWRDQGGIAAARVFTRIWLALFGWWKDDLELPPELMFFPKWVPL  
 NIYDFGCWARQTI VPLTIVSAKRVPVPAPFALDELHTDPDHPNPPKLPAPTSWDGLFQRLDKG  
 LHLYHKVAPRPLRRVAMNLAARWII ERQENDGCWGGIQPPAVYSVIALHLLGYDLDPVPMKAG  
 LASLDRFAVRREDGARMIEACQSPVWDTCLATIALADAGLRPDHPALVKAADWMLAEITRPG  
 DWSVRKPELAGGGWAFEFHNDNYPIDDDTAEVVLAALRRVRHPPARLQAAIDRGVVRNLGM  
 QSRNGAWGAFDADNTSPFPNRLPFCDFGEVIDPPSADVTGHVEMLAVEGLASHPRTRREGIE  
 WLLAEQEA CGAWFGRWGNVYVGTGVSVPALITAGLPAGHPAIRRAVAVLESVQNDGGWG  
 EDLRSYQEEKWI GHGESTASQTAWALLALLAAGRDRTRPVARGVTWLTEAQADGSWDEPY  
 FTGTGFPWDFSINYHLYRQVFLPTALGRYVHGDPFADRAMAAEGA

&gt;seq\_ID 121

MQTQNRVTSTQKVELSNLTKAIIASQNYIMSRQYPEGYWGELESNITLTAETILLHKIWKTDKT  
 RPFHKVETYLRRQNEQGGWELFYDGGELSTSV EAYMALRLLGVTPEDPALIRAKDFILSQG  
 GISKTRIFTKPHLALIGCYDWKGIPSI PPWIMLFPDNFPFTIYEMSSWARESTVPLLIVFDKPIFEI  
 EPAPNLDELAYAGVENVKYALPRNHNSDI FLGLDKLFWTEKNLVPFHKKSQAERWMLN  
 HQQESGDWGGIMPPMVNSLIAPKVLNVDVADPSVQRGF EADRFSEEBEDTYRVQACVSPVWD  
 TAWVIRALVDSGLKPDHPSLVKAGEWLLDKQILEYGDWAIKKNQKPGGWAFEFINRFYPDL  
 DSAVVVMALNGIKLPDENCKKAAINRCLWMA TMQCKPGGWAAPVDNDQAWINEI PYGDLK  
 AMIDPNTADV TARVLEMVGSCLKMDENRVQKALFYLEKEQESDGSWFGRWGVNYI YGTSV  
 LSALAVIAPNTHKPMQEKAVNWLISQCNEDGGWGETCWSYNDPSLKGTVSTASQTAWALIG  
 LLDAGEALETLATDAIKRGINYLDTQTPDGTWEAEFTGTGFPCHFYIRYHLYRHYFPLIALGR  
 YWKIGLKNLKG

&gt;seq\_ID 120

MQTQNRVTSTQKVELSNLTKAIIASQNYILSRQYPEGYWGELESNITLTAETVLLHKIWKTDKT  
 RPFHKVETYLRRQNEQGGWELFYDGGELSTSV EAYMALRLLGVTPEDPALIRAKDFILSKG  
 GISKTRIFTKPHLALIGCYDWKGIPSI PPWIMLFPDNFPFTIYEMSSWARESTVPLLIVFDKPIFEI  
 EPAPNLDELAYAGVENVKYALPRNHNSDI FLGLDKLFWTEKNLVPFHKKSQAERWMLN  
 HQQESGDWGGIMPPMVNSLIAPKVLNVDVADPSVQRGF EADRFSEEBEDTYRVQACVSPVWD  
 TAWVIRALVDSGLKPDHPSLVKAGEWLLDKQILEYGDWAIKKNQKPGGWAFEFINRFYPDL  
 DSAVVVMALNGIKLPDENCKKAAINRCLWMA TMQCKPGGWAAPVDNDQAWINEI PYGDLK  
 AMIDPNTADV TARVLEMVGSCLKMDENRVQKALFYLEKEQESDGSWFGRWGVNYI YGTSV  
 LSALAVIAPNTHKPMQEKAVNWLISQCNEDGGWGETCWSYNDPSLKGTVSTASQTAWAII GL  
 LDAGEALETLATDAIKRGINYLDTQTPDGTWEAEFTGTGFPCHFYIRYHLYRHYFPLIALGRY  
 WKIGLKTSPVLIPLN

&gt;seq\_ID 132

MFQSDRPPVTLVMNDMRGPD MNVSDTVSVTRESIPTQTSAGDATARDLTA AVGSELTRALR  
 LATDHLALQDGTGWWKFDLENTSMDAEDLLREYLGIRTEVTAASARFIRSRQSDDGSWP  
 QYFGGPGELSTTVESYIALRLAGDDASAPHMLSAATWVRDHGGVPATRVFTRIWLALFGWWR  
 WEDLPALPEIMLLPRRAPLNIYSPGWARQTLVSLTVVSALRVPVPAPFDLDELYPDGPASAW  
 SGAGPSNVLERI STRFTAKEIFLGI DRLLHYHRRPVRSMRNHALRAAERWII ARQEA DGCFCGGI  
 QPPAVYSI IALRLLGYELDHPVLKALRALDDYSVTL PDGSRMVEASQSPVWDTALAVNALADA  
 GATAAIAPDHPALVRAAGWLLGQEVRRRGD WAVNHPDVPASGWAFEFENDTYPD TDDTAE  
 VLLALRRVRHPARDELDAERRAVAWL PGLQS DSGGWAGYADANTSTI PYQIPFADFGAL TDP  
 PSADVT AHVVVLLAEAGLGGDDRTRRGV DWLLDHQEA DGSWFGRWGVNYVYGTGSPMPAL  
 RAAGLEPSHPAMRAGADWLLTHQNADGGWGEDLSYTDPEWSGRGESTASQTAWAMLALL  
 TVGDQPEVSGALARGARWLADHQRPDGSWDEDDQFTGTGFPGDFYINYHG YRLLWPIMALGR  
 YLRG

&gt;seq\_ID 118

MLTYKEYRRSVTEIAMQTRDRQTKPALSLNDAITASQNYLLSLQYPOQYWWAELESNITLTAET  
 TVLLHKIWTGDKTRPLHKVEAYLRQQREQGGWELFYDGGELSTSV EAYMALRLLGVPQDD  
 PALIRAKDFILSKGGISKTRIFTKPHLALIGCYSWKGI PSIPPWIMLFPNFPFTIYEMASWAREST  
 VPLIIVFNDKPVFAVDPIFNLDLDELAYAGIENVKYELPKNNNWGDI FLGLDKVFKFAEQVDLVPPRK  
 KGLQAAERWMLNHQJQETGDWGGIMPPMVNSLLAFRVLNVDVNDPSVQRGF EADRFSEEBED  
 TYRVQACVSPVWDTAWCVRAL TNSGLPKDHPSLVKAGKWLLEKQCLEYGDWAVKNTGKPG  
 GWAFEFTHRFYDIDDSAVVVMALNGIKLPDEARKQAAINRCVKWIEITMCKEGGWAAPVD  
 NDQAWLNEVPYDGLKAMIDPNTADV TARVLEMVGSCLDEI SSKRLNKALNYLYKEQEKDGSW  
 FGRWGVNYI YGTSVLSALAVINPEKHQPQIEQGINWLLSQCNDGGWGETCWSYNDNSNLKG  
 KGISTASQTAWALI GLLDAGEALNHFE TDSIQRGISYLLNTQTEBGTWESEFTGTGFPCHFYIR  
 YHFYRHYFPLIALGRYQNLSSSEFGIRNSEL

&gt;seq\_ID 230

MTATTDGSGGLRGGAA TAGETTSTSAARTTEPGTDLREAAAARAERAVEHLLARQDAEGWW  
 KGDLETNVMDAEDLLLRQFLGIQDPATV GASARFIRGQQRDDGTWATFYGGPGELSTV EAY  
 VALRLAGDRPDPHMQRAASVWR SRGGIAASRVFTRIWLALFGWWKEDLELPPELIFL PK  
 WFLPNIYDFGCWARQTI VPLTIVSAKRVPVPAPFALDELHTDPALPNPKRLAPASAWDGFQ  
 RADKALHAYHKVAPRRLRRAAMNAAARWII ERQENDGCWGGIQPPAVYSVIALHLLGYDLEHP  
 VMRAGLES LDRFAVHHEEGLPVRMIEACQSPVWDTCLATIALADAGLPADHPALVKAADWML  
 SEQIVRPGDWSVRRPGLPGGWAFEFHNDNYPIDDDTAEVVLAALRRVRHPPARLQAAIDRGVVRNLGM  
 TRWNLGMQSRDGA WAFDADNTSPFPNRLPFCDFGEVIDPPSADVT AHVVVLLAEHGMADHP  
 RTRRGVRRLLAHQEA NGAWFGRWGNVYVGTGAVVPALTAAGLP GSHPAIRRAVAVLESVQ

## Enzyme Sequences

NEDGGWGEDLRSYREEKSI GRGVSTASQTGWALLLALLAAGERESKAVERGV AHLAQTQAPD  
GSWDEPYFTGTGFPWDFSI NYHLRYQVFPALTALGRYVHGEKLPGRAGAREGR

>seq\_ID 234

MHEGEAMTATTDGSGTAATPPATTASAPLHLSPEARETHEATARATRAVDFLLARQSDGEGW  
WKGDLATNVTMDAEDLLLRQFLGIRDEATTRAALFIRGEQEDGTWNTFYGGPGLSATIEG  
YVALRLAGDSPEAPHMRKASAFVRAQGGVARARVFRITWALFGWKKWEDLPEMPELMMFF  
PKWAPLNIYDFGCWARQTI VPLTVVCAQRVPRPAPFALEELHTDPADPDPAQPAPPVVSNDNV  
FHKLDKLLHGYYRIAPRRVREAAAMRAATWIVERQENDGCWGGIQQPAPVYSIMALNLLGYDL  
HPVLRAGLASLDRFAVWREDGARMIEACQSPVWDTCLATVALADAGVPADHPQMIKAADWML  
AEQIVRPGDWVVRPDLPPGGWAFEFHNDNYPDI DDTAEVVLAARRVAHPDTRVDKAVRA  
VDWNVGMQSKNGAWGAFDADNTSPPFNRLPFSDFGEVIDPPSADVTAHVVEMLAEEGLAHH  
PRTRRGIEWLKNQEGNGSWFGRWGVNYVYGTGAVVPALVAAGLPASHPAIRRSVSWLGGQV  
QNEGGWGEDLRSYQDSAWHGRGHS TASQTAWALLLALLAAGERETEQRVRRGIAYLVETQTE  
DGTWDEPWFPTGTGFPWDFSI NYHLRYQVFPALTALGRYLNGTGPGEN

>seq\_ID 123

MQTRDRQTHKPALS LNDAI TASQNYLLSLQYPPQGYWVAELESNITLTAETVLLHKIWGTDKTRP  
LHKVEAYLRQQQREHGGWELFYDGGGEISTSV EAYMALRLLGVPNDPALIRAKNFII SQGGIS  
KTRIFTKPHLALIGCYSWKGIPSI PIPWIMLPFNSFPFTIYEMASWARESTVPLIIVFNDKPVFAIDPI  
FNLDELAYAEGIENVKYELPKNNWGD LFLGLDKVFKLAEQVDLVPRKQGLQAAERWMLDHQ  
QETGDWGGIMPVMSLLAFRVLNYDVADPSVQRGFEAIDRFSIEENDTYRVQACVSPVWDT  
AWCIRALTD SGLPKDHFSLVKAGKWLLEKQVLEYGDWAVKNTKPKPGGWAFFETNRFPDID  
DSATVVMALNGIKLPDEALKQAAINRCLKWIETMQCKAGGWAAFDVNDQAWLNEIPYGLDKA  
MIDPNTADVTRAVVEMVGS CDLEMSDRLNKALDYL YEEQEKDGSWFGWGVNYIYGTSGVL  
SALAVINPKQHSQIEQGMNLLSCQNEGGWGETCWSYNDLSLKGKGVSTPSQTAWALIGL  
LDAGEVLNHFETDSIERGINYLLNTQTEEGTWESEFTGTGFPCHFYIRYHFRHYFPPLIALGRY  
QQMLGS

>seq\_ID 10

MTQASVREDAKALDRAVDYLLSLQDEKGFWKGELETNVTIEAEDLLREFLGIPTDITAETAR  
WIRAKQRS DGTWATFYDGPDLSTSV EAYVALKLAGDDPAAPHMEKAAAYIRGAGGVERTRV  
FTRLWALFLGLWPDDLP LPEMIFLPSWFP LNIYDWGCWARQTVVPLTIVSALRVPRIPLSI  
DEIRTAGPPPRDPAWTIRGPFQRLDDLLRQYRRVADHGPALFRRLAMRRAAEWIIARQBAD  
GSWGGIQPPHWVSLIALHLLGYPLDHPVLRGLDGLNGFTIREETADGAVRRL EACQSPVWDT  
ALAVTALRDAGLPADHPRVQAAARWLVGEEVVRVAGDWAVRRPGLPPGGWAFEFANDNYPDT  
DDTAEVVLAARRVLEDADQQALEAAVRRATWVIGMQSTDDGGWAFDADNTRELVLRLPFC  
DFGAVIDPPSADVAHI VEMLAALGMRDHPATVAGVRWLLAHQEPDGSWFGWGANHI YGTG  
AVVPALIAAGVSPDTPPIRRAIRWLEEHQNPDDGGWGEDLRSYTDPALVWRGVSTASQTAWA  
LLALLAAGEEASPAVDRGVRWLVTTQQPDGGWDEPHYTGTGFPDFYINHYLRYLVPFISALG  
RYVNR

>seq\_ID 233

MRRRRSPRGPAGPEADYGPASAPDRLRGDAARGDAARRVQDATARAIRNLLGRQDPAG  
WKGDLNVTNVTMDAEDLLLRQFLGIRDEAVTQAAALFIRREQREDGTWATPHGGPPELSATIE  
AYVALRLAGDAPDAPHMATASAWIRAHGGLAAARVFRITWALFGWWDWENLPELPELVLLP  
PWWPLNIYDFGCWARQTI VPLTVVSAMRVPVRPAPFALDELHTDARVVPVPRRMAPPTTWNGA  
FQWMDRALHVYRRFAPRRLREAAASAGRWII ERQENDGCWGGIQQPAPVYSVIALHLLGYDL  
GHPVMRAGLES LDRFAVWREDGSRMIEACQSPVWDTCLAAIALADAGVPDHPALVKAADW  
MLGEEIVRTGDAVAVRRPGLAPGGWAFEFHNDTYPDI DDTAEVVLAARRIRHPDARVVAAR  
GVSWNLGMQSRGGAWGAFDADNTSPPFNRLPFCDFGEVIDPPSADVTAHVVEMLAEEGRAA  
DPRTRRGIAWLLAEQEPGEPWFGWGTNYVYGTGAVVPALTAAGLSPGHPAIRRAVLWLESV  
QNPDDGGWGEDLRSYQDRAWAGKGES TPSQTAWALLMALLSAGERDAKTVERGIAYLVETQTA  
DGGWDEPHTGTGFPWDFSI NYHLRYHVPALTALGRYLYGEPFGHGRHIGALHDRTGVPA  
EGV

>seq\_ID 116

MQTQDRLTQKQPLSLKDAI TASQNYLLSLQYPPQGYWVAELESNITLTAETVLLHKIWGTDKTRP  
LHKVEAYLRQQQREHGGWELFYDGGGEISTSV EAYMALRLLGVPQDDPALIRAKDFIISKGGIS  
KTRIFTKPHLALIGCYDWKGI PIPWIMLPDPSFPFTIYEMASWARESTVPLIIVFNDKPVFVSDP  
VFNLDELAYAEGIENVKYELPKNNWGDIFLGLIDQVFKFAEQVDLVPPRKEGLKAAEKWILNHQ  
QETGDWGGIMPVMSLLAFRVLNYDVNDPSVKLGFEAIDRFSIEEDDTRYLQACVSPWDTA  
WCVRALTD SGLKDHFSLVKAGKWLLEKQVMEYGDWAVKNTKAGKPGGWAFFETNRFPDLD  
DSATVVMALNGIKLPDEARKQAAINRCLQWIE TMQCKEKGWAAFDLNNQAWLNEVYPYGLK  
AMIDPNTADVTRAVVEMVGLS CDLEIESDRLNKS LNLYLKEQEKDGSWFGWGVNYIYGTSGVL  
SALAVINPEKHKTQMEQGINWLLSCQNKDGGWGETCRSYNDPSLKGKGVSTPSQTAWALI GL  
LDAGEALNKFTDAIERGVNYLLDTQTEEGTWESEFTGTGFPCHFYIRYHFRHYFPPLIALGRY  
YQNLSSSEFGVRS

>seq\_ID 124

MQIRATVDTAKLEKIAAASQEHLLSTQYPEGYWVAELESNVTMTAEVLLHKIWKTDGTRPMH  
KAEKYLRSQREHGGWELFYDGGDLSTSV EYTYTALRLLGVPASDPALLKAKDFILRRGGISKT  
RIFTKLHLALIGCYDWRGLPSLPPWVMLLPENFPFTIYELSSWARGSTVPLIIVMDRKPVFSVNP  
QINVDELAYAEGRDRVKFELPRKGDWTDLFI EL DGLFKFTEQNNLVPPFREGLRAAERWVLERQ  
EATGDWGGIIPAMLNSLLALRALGYHPADPYVRRGMAAVDRFAIETADTYRVQPCVSPVWDTA  
LVMRGLIDSGLPADHPAIVKAGEWLLLEKQI LAYGDWAVKNTKQPGGWAFFEFENRFYDPVDDSD  
AVVVMALQAAQLPDEDLKQAI ERVCVKIATMQCKPGGWAFFDVNDQDWNQIPYGLDKA

-continued

Enzyme Sequences

MIDPNTADVTRVLEMIGRSGVTTGEASVERALAYLRREQEVEGCFWRGWVNYIYGTSGVL
AALALIAPKSDHAMIQRGADWLVRQCNADGGWGETCRSYNDPHLKGQGPSTASQTAWALIGL
LAAGEATGEFAWGAIDRGINYLLATQQQDGRWDEDWFTGTGFPGHFYLYKHYLQQHFLTLAL
GRYSSLTGLKQELKIPLQLKSKPEVVMIEDSDLLSDEDAT

>seq\_ID 119
MQIQDRNSSPQVTEVLNQKDAIAASQDYLMISIQYPEGYWWAELESNVTITAEVLLHKIWGTD
KTRPLHKVETYLRRQQRHEGGWELFYDGGDLSTVEAYMALRLLGVSIIDDPALIRGREFILKR
GGISKSRIFTKLHLALIGCYDWRGIPSPWPIMLLPENFPFTIYEMSSWARSSTVPLLIIVFDKPKVY
CCDPTINLDELYSEGIENVKYDLPKTGDWTDIFVWLDGVFKFAQDYNLVLPRQESLQAAERWV
LERQEDSGDWGGIIIPAMLNSLLALRALNYEAVDPVHRGLQSDVNFATIEDTYHVQPCISPVW
DTAWAIRALVESGLKADDPRLVKGAQWLLDKQILDYGDWAVKKNQGTGGWAFEFDRNRWYP
DLDDSAVVVMALDQVKMPNEDLKNGAI RRCVRWMATMQCKDGGWGFAPDLNDQNLNLFPLP
YADLKAMIDPNTADVTARVLEMLGTCLIMDSNRVQKAIAYLEKEQEPDGSWFGRWGVNYIYG
TSGVLSALAVIAPETHQKELKGAAWLVGCQNDGGWGETCFSYNDSSLKGGKDSASQTA
WGLIGLLAAGEATGEFFKTAIERGVNLLKTQREDGTWEDENYFTGTGFPCHFYLYKHYLYQYFP
LIALSRYQRLLT

>seq\_ID 9
MSVSERAQPGGNPIPGSTSQS AVKFRIDALEDVKRAIAGAKDRVFAQQSKDGNWCGELEA
DSMLEADYIFAHTLLGTGAGKMKRALTEMLRYQNEGDSWSIYPPGGPNTISLTVKCYFSAKLM
GMTADNPLLVKAREWILAHGGVVECENTFTKIYLCFLGQYEVDAVPAIPEI VLPFNWFYFNIYEIS
SWSRAILVPLSIAYAKKPKKIPPEQGIIDELFVGGREKANLHLRWDKSNLLSWRNFALALDRVTH
WFERVHIRPLRSIALKKAEKWMLARFEMSDGLGAIYPAMLNAIALRCLGYSLDDPQVLRAMDE
FEKLGIDEPETGAEYAEPTFRMQPCVSPVWDTAQAVFALGEAGVPRNDPRMQKAADWLLSKE
VRHKGDMAMKVRNAQPGGWYEFNNEFYPDVDDSAQVLLALNKNVDNPRERYQYDVCQRAID
WIFAMQCRNGGASFDKNDTKMIQYVVPFADHNAMLDPPTVDITGRILEMLATYGYTRKDRRV
EKAIKFIYDEQEPDGSWFGRWGVNLYGTFLVLRGLEAIGVWNHEPQIQQAARWIRSVQNDAG
GWGETCGSYDDPNTRGVGSPSTPSQTAWAILGLLSAGDDRSDSVAKGIKWLALAHQKPDGGWD
ESTGSGSKHQALYTGTFPRVYLYAYHQYRDYFPLLALTYEKAMERGE

>seq\_ID 217
MTEEVLRQTAAPAEVLAAREHLLSLQHERGWKGELETNVTMDVEDLLLRFLGILTTAETE
QAARWIRSRQRADGTWAFHGGPGLDSTTVEAYVGLKLAGDDVDSEHMAAARAWILERRGGIE
ETRVFTRIWLAGLFGESWDDLPAMPPELVLPVLPWVPLNLADWGCWARQTI VPLTVVCTLRPR
RDLGVGLAELRSGRRRKVPSPSWAGAFQVLDGALHGYQRHPLRGLREHAMRRAAEWIVAR
QEADGSGWGGIQPPVYSLALHLGLYPLDHPVLRQGLAGLERFLIREETPEGTVRRLEACQSP
VWDTVLSMQLRDAAGLADHPALRRAADFVLAEEIRVKGDWSVRRPDLAPGGWAFEFNDNG
YPDIDDTAEVVLALNRVDHERPGAVNAIDRGVRWMSGMSADGGWGFADADNTRELVNEL
PFCDPGAVIDPPSADVTAHVVEALCVLGRGDGEAVRRGVRLLDHQELDGSWFGRWGANHV
YGTGAAPALVRAGLRDRDHLALRRRAVRWLEVHQNDGGWGEDLRSYDDPVVWVGRGRSTAS
QTAWALLALLAVLDHDTAVRRGVGFLAETQRPDGTWDEPQFTGTGFPDGFYINHYLRLVFP
VTALGRYEQARREQSGGSG

>seq\_ID 249
MIEKNKVKQSIILASQKHLSSLQETEGYWWGQLESNVTITAEIILLHKIWQTDKKIPLNKAKNYLIS
QQREHGGWELFYDGGDLSTSEAYMALRLLGVSRTPDPIIMIEAQNFIIKKGGISCSRIFTKLHLAL
IGCYSWQGIPISSIMLLPEDFPFTIYEMSSWARSSTVPLLIIVFDKPKIFSVNPTINLDELYAEGI
NNASFELPRKYDLDLFLGLDKAPKAENLMLPQQEGLKAAEKWI LERQEVTDGWGGIIPAM
LNSMLALKCLEYDVADPVVVRGLEAIDRFAIENEDSYRVQACVSPVWDTAWVIRSLVDSGISPS
HPAMVKAGQWLLQQQILDYGDWVFKNFKPGKPGGWAFFEMNRFYPDIDTAVVVMALDVLVEL
PDEDLKGKAIARGMHWIASMQCEAGGWAADFVDDNNDQDLNATPYGDLKAMIDPNTADVTR
VLEMVGCCLAMDSWRVVRKRGIDFLVREQEEGCWFGRWGVNIIYGTSGVILALAVMARESHR
GYIERGASWLVGCQNSDGGWGESCWSYNDPSLKGKGSASQTAWALIGLLAAGEGTGNFA
RDAIDGGVGLVSTQNDGGSWLEDEFTGTGFPGHFYIKYHFSYQYFPLMALGRYESLLSG

>seq\_ID 222
MAVRDRVNPKLEAAIAASQSYLLTQQDETGYWWAELESNVSITSEVLLHKIWGTDRSRPLE
KVETYLRSQQRDHGGWELFYDGGELISVSV EAYMALKLLGVPMEDPAMVRRARQFILLEHGGISR
TRVFTKLHLALIGCYEWRGIPSLPPWVMLLPEQFPFTIYEMSSWARGSTVPLLIIVMDREPVIYAV
EAGFNLDELYEGRHRAQFDLPLSNEWTDAPFIYLDGLFKFAESTNLVPPREEGIRAAERWILER
QEATGDWGGIIPAMLNSLLGLKALDYVHDPIIERGMAALDAFALETEDQYWIQPCISPVWDTA
LVVRGLAESGLAPDHPALVKAGEWLLNKQILDYGDWVSKNPGGLPGGWAFFEDNRFYDPVDD
TAVVVMALNEVQLPDEQAKDAIARAVNWIATMQRCPGGWAAFDDINNDQDLNALPYGDLKA
MIDPNTADVTRVLEMIGRCHQTTGKNSVDRALRYLREQEPEGCFWRGWVNYIYGTSGVL
AALALIDPQGWQSIQQAAAWLVSCQNTDGGWGETCASYDNPCLKGQGPSTASQTAWAIMG
LLSAGEATSVYAEAAIERGVNLTQKMDGTWDEDFYFTGTGFPGHFYLYKHYLQQHFLTLAL
GRYQAMLQOKS

>seq\_ID 186
MRTQDRVQVNSIAEAAIAASQYLLSLQNPAGYWWAELESNVTITAEVLLHKIWGTDKTRPLHK
VEAYLRSQQKHGGWELFYDGGELISVSV EAYMALKLLGVPATDPAMIQARDFILQRGGISKT
RIFTKPHLALIGCYNWRGLPSLPWVMLLPNQFPVNIYEMSSWARSSTVPLLIIVFDKPKVYQVN
PTITLDELYAEGVENVRYELPRSGDWTDLFLTLDEGFKLAESFNIPFREEGIIKAAEKWI IERQEA
TGDWGGIIPAMLNSMLALRSLGYDTNDPIVERGLQALDNFAIETVDCYRVQPCVSPVWDTAWV
RALIDSGITAPDHPAIVKAGEWLLQKQILDYGDWVKNRQGGKPGWAFEFENRFYDPVDDTAVV
VMALHAALKPNEQLKQKACDRALQVAVSMQCKPGGWAFFDLNDQDLNLSVYGDLMKAMID

## Enzyme Sequences

PNTADVTRARVIEMLGACNLSIDSHNLERALTYLLNEQEAEGCWFGRWGVNIIYGTSGVLSALAL  
INPQKYQRHIQQGATWLVGCQNPDDGGWGETCFSYNDPSLKGQGDSTPSQTAWALIGLIAAGE  
ATGNFAHDVIERGINHLVSTQQPDGSGWFEAYFTGTGFPCHFYLYKHYYQQYFPLIALGRYQAIK  
SL

>seq\_ID 153

MQVQPRIIEKKHLSDAIEASQAYLLARQYSPGYWAAELESNVSMTEAVVLLHKIWRDTGRPLA  
KATAHLAEQRAHGGWELFYDGGDLNNTSEAYMALKLLGLTADHPALARARAFILAKGGISRA  
RIFTKIHLLALIGCYDWRGVPSIPPWVMLLPEAFPVNIYEMSSWARGSTVPLLIIVFDRKPVFAVEP  
AITLDELVEGRAQARFDLPRSSSDWWANLFDVLDWGFKLAESLGAVPLREGLKAAERWVLE  
RQEATGDWGGIIIPAMLNSLLALRCLDYDPHPDPPVVRGMAAVDRFAIETESTYRLQPCVSPVWD  
TALTMRALVDSGLPPDHPALAAAGTWLLKKQILDYGDWAVKNRTGPPGGWAFEDNRFYPDV  
DDTAVVVMALDAVRLADETAKGQAIARAVCWVASMQRGGGWAAFDIDNDAHWLSLPHYAD  
LKAMIDPNTADVTRARVIEMLGACNLSIDSHNLERALTYLLNEQEAEGCWFGRWGVNIIYGTSGVLSALAL  
LGLIDASRVARFSDSSALERGLAYLVETQKADGSDWDEPYFTGTGFPCHFYLYKHYYQQYFPLIALGRYQAIK  
ALGRYRRLLS

>seq\_ID 122

MQIQARNISTKVTVEFVSKVKEAIAASQQYLLSIQYPEGYWAAELESNVITAEAVLHKKIWTDT  
TRPLHKVETYLRRQREHGGWELFYDGGDLNNTSEAYMALRLLGVSASDPALVRAKAFILSR  
GGISKSRIFTMHLALIGCYDWRGVPSIPPWVMLLPEAFPVNIYEMSSWARGSTVPLLIIVFDRKPVFAVEP  
VYQCGITLDELVEGRAQARFDLPRSSSDWWANLFDVLDWGFKLAESLGAVPLREGLKAAERWVLE  
LERQEDTGDWGGIIIPAMLNSLLALRCLDYDPHPDPPVVRGMAAVDRFAIETESTYRLQPCVSPVWD  
DTAVVVMALDAVRLADETAKGQAIARAVCWVASMQRGGGWAAFDIDNDAHWLSLPHYAD  
LDDSAVVVMALDAVRLADETAKGQAIARAVCWVASMQRGGGWAAFDIDNDAHWLSLPHYAD  
PYADLKAMIDPNTADVTRARVIEMLGACNLSIDSHNLERALTYLLNEQEAEGCWFGRWGVNIIYGTSGVLSALAL  
LGLIDASRVARFSDSSALERGLAYLVETQKADGSDWDEPYFTGTGFPCHFYLYKHYYQQYFPLIALGRYQAIK  
ALGRYRRLLS

>seq\_ID 129

MSLTSDDPSPAAPTAEKSPKRPTIPVPATADAYGISRSSPPLPAATGRPQAAGPASAGVATARAR  
DHLALQSEEGWKKGDLETNVTMDAEDLFMKQFLGIRGDDTEQTRAWIRSQQLADGGWPT  
FYGGPADLSTTIEAYIALRLAGDAVDAPHMARAAELVRAQGGVAASRVFTRIWLAALGQSWD  
DVPVIPPPELIFLPSWIPLVNVDYFACWARQTIIVALTIVGSLRPSHDLGFSIDELKVPAAARKPAALR  
SWEGAFERLDKLLHRYEKRPILKRLTLALRRATEVWVARQEDAGCWWGIIQPPWVYVSMALHL  
MGYPINLHPIVIAAFRGMERYVIRRDTPQGIHQIEACQSPVWDTALAVVALADAGVPGDHPAM  
VKAGRVLVDEEVRVAGDVAWRPELAPGGWAFEDNDFYPDVDDTAEVVLALRRLGAGHV  
APPASRQGRAEAPPVNTVEDADPRLAAAMRAAAARGVDWVGMRSNNGAWGAFDADNVRT  
LTTKIPFCDFGEVVDPPSADVTAHIVEMLADLGRSDHPITQRAVQWLLDNDQEPGGSWFGRWG  
VNHLVGTGAVVPALIGAGVPTDHPAITAAVRWLLEHQSPGGWGEDLRSYTDPAWIRGELTA  
SQTAWALLALAVDPHSLAVKRGVRLCETQRPDGTWDEPYFTGTGFPCHFYLYKHYYQQYFPLIALGRYQAIK  
ALGRYRRLLS

>seq\_ID 164

MHSGRVFLEKENRENRATFHSSPLILVEESLNLKPKVEETIKKAQRVYLLSIQKEDGHVWVGELEF  
VDVTLACDCIHLMHWRGKIDYKQQLRVLKHIIVDRQLPDGGWNIYPPGSPSEVNATVKAYFALKLA  
GFSPDDPLMAKARSTILRLGGIPKCMYTKLGLALLGVYPWDRLPVIPPETILFPNWFPPNIYEISA  
WSRAMLVPLSVIHHFKPTRNLEPEYQLHELFPYGTTEHGKFSWLKKGARYLSKQGLFLACDKFL  
QYWDKTSKPPFRKMAKKAELWLLERI SAGSDGLGAIFFAMHYAIMALIAMGYTEDNPIKKAIA  
DFEGLEVDKNDLRIQPCLSPVWDTAVGLVALAESGVARNAKELKRAAVVLLDREIKIKGD  
WHVRNPHPEPSGWAFAEYNNVYYPDVDDTLMVLLALRLIDIEDKIRKEEVMQRALRWVIFSQCK  
NGGWAAPDKVYKWLLEDIPFADHNAI LDPPCSDITARELFGKMGIKKTERFVQKAIAYLKET  
QENDGSMWRGVNIIYGTWQALRGLQAI GENMNQEWI LRARDWLESQNDGEGWGETP  
ASYDNPQLKGGKPS TASQTAWAVSGIMACGDI FRPSVSRGIKYLCDRQLSDGSAEELTGT  
GPPGVFYLKYDMYRNAWPLLVIGEYHRQYLKAKEQVSYWVDGTIGRKVKKERLPEI

>seq\_ID 20

MRTQDRVQVNSIAEAIASQYLLSLQNPTGYWAAELESNVITAEVLLHKKIWTDTKTRPLHKI  
EAYLRSQQKHGGWELFYDGGGELNNTSEAYMALKLLGVPATDPAMIQARDFILQRRGGSKTR  
IFTKHLALIGCYNWRGLPSLPAWVMLLPEAFPVNIYEMSSWARGSTVPLLIIVFDRKPVFAVEP  
AITLDELVEGRAQARFDLPRSSSDWWANLFDVLDWGFKLAESLGAVPLREGLKAAERWVLE  
RQEATGDWGGIIIPAMLNSLLALRCLDYDPHPDPPVVRGMAAVDRFAIETESTYRLQPCVSPVWD  
TAVVIRALIDSGMAPDHPAIVKAGEWLLQKQIFDYGDWVKNRQGGPAGAWAFEDNRFYPDVDDTAVV  
VMALHAAKLPHPQLKQKACDRALQVWVSMQCKPGGWAAFDIDNDQDNLNAVYPGLDKAMID  
PNTADVTRARVIEMLGACNLSIDSHNLERALTYLLNEQEAEGCWFGRWGVNIIYGTSGVLSALAL  
INPQKYQRHIQQGATWLVGCQNPDDGGWGETCFSYNDPSLKGQGDSTPSQTAWALIGLIAAGE  
ATGNFAHDVIERGINHLVSTQQPDGSGWFEAYFTGTGFPCHFYLYKHYYQQYFPLIALGRYQAIK  
PL

>seq\_ID 185

MQTQDRVQVNSIAEAIASQYLLSLQNPTGYWAAELESNVITAEVLLHKKIWTDTKTRPLHKI  
KVEAYLRQEQHGGWELFYDGGGELNNTSEAYMALRLLGVPATDPAMIQARDFILQRRGGSKTR  
TRIFTKHLALIGCYNWRGLPSLPAWVMLLPEAFPVNIYEMSSWARGSTVPLLIIVFDRKPVFAVEP  
TINLDELVEGRAQARFDLPRSSSDWWANLFDVLDWGFKLAESLGAVPLREGLKAAERWVLE  
RQEATGDWGGIIIPAMLNSLLALRCLDYDRSDPIVERGLQAINFAIETDNSYRVQPCVSPVWDTAVV  
IRALIDSGMAPDHPAIVKAGEWLLQKQIFDYGDWVKNRQGGPAGAWAFEDNRFYPDVDDTAVV  
VMALHAAKLPHPQLKQKACDRALQVWVSMQCKPGGWAAFDIDNDQDNLNAVYPGLDKAMID  
PNTADVTRARVIEMLGACNLSIDSHNLERALTYLLNEQEAEGCWFGRWGVNIIYGTSGVLSALAL  
INPQKYQRHIQQGATWLVGCQNPDDGGWGETCFSYNDPSLKGQGDSTPSQTAWALIGLIAAGE  
ATGNFAHDVIERGINHLVSTQQPDGSGWFEAYFTGTGFPCHFYLYKHYYQQYFPLIALGRYQAIK  
PL

## Enzyme Sequences

VMRALVESGFVDPDHAVVKAGEWLLQKQILDYGDWAVKNRQKPGAWAFEFENRFYDPVDD  
SAVVVMALHLAKLPNEKIQAAIARAVNWIASMOCCKPGGWAFFDLNDQDNLNSIPYDGLKAM  
IDPNTADVTRVVEMLGACDLSIDSDNLERSLTYLLREQETEGCWFGRWGVNYIYGTSGVLSA  
LALIDPQRHKLSEIRGAAWLLGCQNLDDGGWGETCRSYDDPSLKGKGDSTASQTAWALIGLLAA  
GEATGKLVAKAIEQIGIYLMATQQPDGTWFEANFTGTGFPFCYFYLKYHLYQQYFPLIALGRYQ  
AAIKES

>seq\_ID 244  
MVAASPSVPCPSTEQVRQAI AASRDFLLSEQYADGYWWESELSNVTITAEVVLHKKIWGTAAQ  
RPLEKAKNYLLQQQRDHGGWELLYDGGELSTVSEAYTALRILGVPATDPALVKAKNFIVGRG  
GISKSRIFTKMHLALIGCYDWRGTPSIPPWVMLLNNFFNIYEMSSWARSSTVPLMIVCDQKP  
VYDIAQGLRVDELAYEGMENVQYKLPESGTIWDIFIGLDSLFKLEQAKVVPFREQLALAEKW  
LERQEVSGDWGGIIPAMLNSLLALKVLDYVNDLYVQRGLAAIDNFVETEDSYAIQACVSPVW  
DTAWVVRALAEADLKGDPALVKAGWLLDKQILTYGDWQIKNPHGEPGAWAFEDNNFYD  
IDDTCVMMALQGITLPDEERKQGAINKALQWIATMQCKTGGWAFFDIDNDQDNLNQLPYGDL  
KAMIDPSTADITARVVEMLGACGLTMDSPRVERGLTYLLQEQDGSWFGRWGVNYLYGTSG  
ALSALAIYDAQRFPQIKTAIAWLLSCQNDGGWGETCESYKKNQKQKQGNSTASQTAWALIG  
LLDALKYLPSPGQDAKLTATAIEGGVAFVQGGTPKGTWEEAEYTGTFPCHFYIRYHYRQYFP  
LIALARYSHLQAS

>seq\_ID 109  
MDDRHIQSEITFGKIDGIRERIQQAMDAAKRYLFSKQDPEGFWCGELEADTTLQSDYIVMHTLL  
GTGDPVKMQKAGKQILQHQNPDGGWNIYPDGPNSISAAYKAYFSLKLIHGKPEPEMTKARE  
WILAHGGVTACNTFSKMYLCPFGQYDYDTPAIPPEIVLFPNWFNLYEISSWSRGI LVP LAIC  
YAKKPFKIPDEANIDELFVEGRHANLHLTDWKKPFSWRNPFVLLNMMVHFERVHVRLRKLKLA  
MKRAEKWMLERLEMSDGLGGIYPAI LNSIIALRALGYSTDDPQVIRAMDEFEKLGIEEDDTFRM  
QPCMSPVWDTAYALYALGEAGVPGSDPRMQKAAEWMLKKQVTHKGDWAVKVRNVQPGGW  
YEFNNEFYDPVDDTAQVILSLNHVRTSNERYQDDTVKRALDQWLAQCKNGGWSFKDN  
NKMVFQYIPFADHNAMLDPATVDITGRVLEALSHHGYSKDKVQRAVKFQIQSEQEPDGSWFG  
RWGVNYIYGTMLCRLGLAAVGDVHHEPMVQQAEEWLRMVQNPDDGGWGESVGSYDDPKLRG  
QGPS TASQTAWAVMGLLAANDLRSVTRGIAWLLLENQKPNGSWWEKVI TGTGFPFVFLKY  
TMYAEYFPLIAFAEYLRRLNTPLEDEKVKLGPQA

>seq\_ID 174  
MQIQDKITEIAAKTAKAIELSQNYLLSTQYSEGYWWESELSNVTITSEAILLHKKIWKTDKKRPLDK  
AATYLRQQCCPNGAWELFYDGGDLSTTVEAYMGLRLLGI PANDPALEKAREFILAKGGISKTR  
IFTKMHLALIGCYDQVGPSPAWIMLLPENFPFTIYEMSSWARGSTVPLLI VPDKPKVYKMGFN  
LDELTYEGVNNVYELPKNNNWSDVFLWLDGLFKWAEKTDLVPRQESLKAEEKVIERQED  
TGDWGGIIPAMLNSLLALKALDYDVPDIVARGLKAVDNFAIETDNTYCVQPCVSPVWDTAWVI  
RSLIESGLNPAHPAMI KAGQWLI DQQILDYGDWAIKKNIGTPGGWAFEDNRWYPDLDSDAVV  
VMALIELKMPDENIKTSVMKRAVNWMTMQCKAGGWAFFDIDNDQDNLNSLPYADLKAMIDP  
NTADVTRVLEMLGCTDVKMGENRVKALDYLEKEQEAADGSWFGRWGVNYIYGTSGALSALA  
FLEPNQYRQQKQGANWLS SCQNVDDGGWGETCFSYNNPKFKGQGNSTASQTAWALIGLLAV  
GKVTGNVQREVIKGVNYYLLVTQKENG TWEDYFTGTGFPCHFYLYKHYFYQQYFPLIALGRYR  
ALI

>seq\_ID 130  
MSLTS D P S P A A P K A A K S K R V N I P A P A T P D A Y G I S R S S P P L S G G G V S G G G V S G G G A A T A D G T P  
P T T Q T S V D P D L A A A M T A A N Q A R D H L L G L Q S E E G W W K G D L E T N V T I D A E H L P M K Q F L G I R T E E E  
T E P I A R W V R S Q Q L A D G G W A T Y Y G G P A E L S T T V E A Y I A L R L A G D E P D A P H M A A A A A I R S Q G G V  
A A A R V F T R I W L A T F G E W S W D D V P V L P P E L I F L P S W F P L N V Y D F G C W A R Q T I V A L T I V G S L R P V R  
D L G F S I D E I K V A A P V T P K P A P L H S W E G A F E R L D A I L H R Y B E R R P I K V L R T L A L R R A T E W V V A R Q E  
A D G C W G G I Q P P W I Y S V M A L H L M G Y P L N H P V I A T A F R G M E R Y I I R R E T P E G P T A Q I E A C Q S P V W  
D T A L A V V A L S D A G V P A D H P A M V R A G R W L V D E E V R V A G D W A V R R P A L A P G G W A F E F D N D F Y P  
D T D D T A E V V L A L R R L L G G S H V T P G G T V T P S G S V T P G G T A E L S P A A R D R A S R G L A A V D P Q L A G  
A M R A A A A R G V D W S V G M R S D G A W G A F D A D N V R T L T A K I P F C D F G E V V D P P S A D V T A H I V E M L  
A D L G R S D H P I T R R A V Q W L L D N Q E P G G S W F G R W G I N H V Y G T G A V V P A L I A A G V P A D H P A I T A A V  
R W L L E H Q S P D G W G E D P R S Y D D P A W I G R G E L T A S Q T A W A L L A L A V D P H S K A V K R G R V R W L C  
E T Q R P D G T W D E P Q F T G T G F P G D F Y L N Y H L Y R L V F P L T A L G R Y V T L T G V A T P

>seq\_ID 248  
MPTSLATAIDPKQLQQAIRASQDFLFSQQYAEQYWWESELSNVTMTAEVVLHKKIWGTQORLPL  
AKAEQYLRNHQRDHGGWELFYDGGDLSTVSEAYMGLRLLGVPETDPALVKARQF ILARGGI  
SKTRIFTKLHLALIGCYDWRGIPSLPPWIMLLPEGSPTIYEMSSWARSSTVPLLI VMDRKFVYG  
MDPITLDELAYSSEGRANVWELPRQGDWRDVF IGLDRVFKLEFETLNIHPLREQGLKAAEWWL  
ERQEASGDWGGIIPAMLNSLLALRALDYAVDDPIVQRGMAAVDRFAIETETERYVQPCVSPVW  
DTALVMRAMVDSGVADHPALVKAGEWLLSKQILDYGDWAIKKNKGRPGGWAFFEFENRFYDP  
VDDTAVVVMALHAVTLPNENLKRAIERAVAWIASMOCRCRPGGWAFFDIDNDQDNLNGLPYGD  
LKAMIDPNTADVTRVVEMLGACGLTMDSPRVERGLTYLLQEQDGSWFGRWGVNYLYGTSG  
GVLTALSLVAPRYDRWRIRRAAEWLMQCNADGGWGETCWSYHDP SLKKGKGDSTASQTAW  
AIIIGLLAAGDATGDAYEAIERGIAYLLETQRPDGTWHEDYFTGTGFPCHFYLYKHYFYQQYFPLIALGRYR  
ALGRYARWRNLLAT

>seq\_ID 150  
MAKGILNKFAVIAGTKKAGPPAGEERTVIAPIKEISGKAVHCSQAVKKAEEYLLALQNPGEYVWF  
BLEADVITPSBYIMLQRLGRIEISPELGKRLENYLLDRQLPDGGWPLYAEDGFANISATVKAYLA  
LKVLGHPQAPHMIRARLMVLSLGAARCNVFRILLALFGQIPWHTPPAMPVEIVLLPQWFFF

## Enzyme Sequences

HLSKVSYWSRTVIVPMLLLYAKQPVCLRLPEEGIPPELSTPPDKLRHLDFGQPGYWRKNAFIIFD  
 RLLKRFNRFPISALHRKAIABEQWTRSHMQSGGIGAIFFPAMAYAVMALRVLCGEGDDPDYIR  
 GLQAIIDLLQHRTPQEADPPRTDGTICIDSGMSAAFALTPSAHAAADGTGSSSICQPCNSPIWD  
 TCLSLSALMEAGMPASHPAATQAVEWLLSQQILSPGDWLSKVPDLEGGGWAFQFENTLYPDL  
 DDTSKVIMSLLRAGALENERYRDRIRAGVNWVLMQSSDGGWAFFIDNNYHYLNDIPFADHG  
 ALLDPSTSDLTGRCEILLSMVGFDRTFPPIARGIGFLRSEQEENGAWFGRWGVNYIYGTWSVLS  
 GLRQAGEDMQQPYIRKAVGWLASCONHDGGWGETCYSYDDPSLAGKASTPSQTAWSLLG  
 LMAAGEVNSLAVRRGVRVLLDHQNGWGTWEEKHFTGTGFPRVYLYRHYGRHFPLWALGV  
 YSRLSSGQKACQDERRHASPGDLHLPWLERIKKR

>seq\_ID 128

MPDLELRDVRADGRHAPNLGRDITLSPSAPTGEPAAPASTPAAVATPTPTPTTAPAPAPAPE  
 NALRETVQRAAEHLRLQDPRGWKFDLETNPTMDAEDLLREYLGI RTVEQTEATAKHIRSR  
 RLDDGSDPTTYFGGPELSTTVACYIALRLAGDSPDEPLRRSAAMIRERGGIPATRVFTRIWLA  
 LFGWWRWEDLVPVPEIMFLPPRAPLSIYSPASWARQITVPLTIVSAARPQCPAPFDLAELEDD  
 EYVPAQSHGAAQSPDTRSPAGGRTLRGAMRLLGGDRPNTAKVFFRGLDAALHRYHRHPIGPL  
 RRHALRTAERWIARQEQADGCFGGIQPPAVYSIIALRLGYLDLHPVLAALRLSDAYTLHREDG  
 SRMI EASQSPIWDTALAVLADAGIDAPADVDVAPALPTQRVATGAPAPSAPVPTALERAADW  
 LLGQEIQRHRGDWAI THPGVAPGWAFEFNDNDYPTDDTAEVVLALHRLNRLRRLRHPTNTR  
 IDAALERSTAWLFAQSRDGGWGA YSDNASTLVYQIPFADFGALTDPSADVTAHVVELLCE  
 TGRIRDPRTRLRGVDWLLRNQEQADGSWYGRWGVNYVYGTGSLPALQAAGLPPHTPAMVAGA  
 RWLLSRQNSDGGWGEDIRS YGDPAWSGRGLSTPSQTAWAMLGLLATDHGGVHADALAAAA  
 RWLTEQQRPDGGWDEEMFTGTGFPGFYLYNYHYGRVLPVFMALGRYLHRSRQHPSD

>seq\_ID 131

MSLTSQSSAAPTAAQSPKIPNPSVARPSADAGSFETAGAVRTDVSIDSVSTGTPVDPVVG  
 AMRRGRDHLLSLQAEEGWKGELTNVTMDAEDMLRQFLGILTPSTATETGRWIRSQQLSQ  
 GGWATFYGGPSDLS TTI EAYVALRLAGDDPDAPHMRSAAEWVRSAGGIAASRVFTRIWLA LAF  
 EWSWDDVPLPAEMTFLPPWFLNIYDFACWARQTVVALTIVGSLRVPVRSFGFTDELRLVQAP  
 KATKAPLRSWAGAFERLDSVLHRYEKRPFPQPLRRLALRRAAEWVIARQEQADGCWGGIQPPMV  
 YSIMALHLMGYPLNHPVISMALFRALDRFTIREETPEGTVRRIEACQSPVWDTALAVVALADAGL  
 GGDHPAMVRAGRWLADAEVVRVAGDWAVRRPTLAPGGWAFEFNDNFYDPVDDTAEVVIARR  
 LLGDGHGPVDSHSDGSGPSAATAASAAAEEAAVAAAGTIAAADPELAARLRAAAERGVDSV  
 GMRSNGAWAAFDADNVRTLVRKIPFCDFGEVVDPPSADVTAHMVEMLALLGRSDHPITQRG  
 VRWLLDNQEAAGGSWFRWGVNHVYGTGAVV PALISAGVDAEHPAIVS SMHWLVEHQTPEGG  
 WGEDLRSYRDEWIRGRGEPTASQTAWALLALLAEPASGTAEWAEVERGVRLCDTQRPDG  
 TWDEPQGTGTFPWFDSINYHLYRLVFPVTALGRVYTLTGRSTS

>seq\_ID 242

MSISALQTDRLSQTTLTQSVVAAQQHLLSIQNPPEGYWANLESNASITAEVVLHKKIWTLSQ  
 LAKLENYLRAQKTHGGWELYNWDGELSTVSEAYMGLRLLGVPSDPAVVKAKQFILLHRGG  
 VSKTRIFTKFHLALIGCYRWQGLPSLPAAWVMQLESPPFFSIYELSSWARGSTVPLLIVDFDKKPVY  
 PLQSPSTLDELFTESAENVRWELEEKGDWSDAFLWLDKAFKLAESVDLVPFREESIRKAEKWV  
 LERQEPSGDWGGIIPAMLSMLALRALGYSVSDPVVRRGFQAI DNFMVSETECWAQPCISPV  
 WDTGLAVRSLTDSGLSPNHPALVKAGEWLLDKQILSYGDWVSKNPQGGWAFEFENSFY  
 PDVDDTAVVAMALQDITLNEPLKRRARIARAVRWIATMQCKTGGWAFFIDNNQDWLNDIY  
 DLRAMIDPSTADITGRVEMHGRFAADLDLANSYAADLSPYRLSRGLNYLKEQELDGSWFR  
 WGVNYIYGTGQALSALALIPERCRIQIERGIWVSVQNAADGGWGETCESYKDKSLKGGKILST  
 ASQTAWALLGLLDVDFCLDPAKIAVDRGIQYLVSTQSEGTWQESFTGTGFPQHFYLYRRLY  
 CHYFPLMALGRYQRVINSAGI

>seq\_ID 143

MAKGILNKFAVIAGNKAGLTAEEECTVVAPIKEVSGKAVHCRQAVKMAEYLLALQNPPEGYW  
 VFLEADVITPSEYIMLQRFLGREISPELRMLENYLLDRQLPDGGWPLVAVDGFANISATVKAY  
 LALKVLGHSPQAPHMIRARIMVLSLGGAAACNVFTRILLALFGQLPWHTPPAMPVEIVLLPQRF  
 FHLSKVSYWSRTVIVPMLLLYAKQPVCLRLPEEGIPPELSTPPDKLRNLDGFGQSGRWRKNAFI I I  
 DRLLKRFNRFPISALHRKAIABEQWTRSHMQSGGIGAIFFPAMAYAVMALRVLCGEGDDPDY  
 VRGMQAIDLLQHRTPQEADSPRTGGPCIDSGTSAFAFDPSPHAAADGRGNSSICQPCNSPI  
 WDTCLSLSALMEAGMPASHPAATQAVEWLLSQQIFSPGDWLSKVPDLEGGGWAFQFENTLY  
 PDLDDT SKVIMSLLRAGALENGLYRDRVARGVNWVLMQSSDGGWAFFIDNNYHYLNDIPF  
 ADHGALLDPSTSDLTGRCEILLSMVGFDRTFPPIAQGI GFLRSKQEGSGAWFGRWGVNYIYGT  
 WSVLSGLRQAGEDMQQPYIRAVGWLTSCONHDGGWGETCYSYDDPSLAGQESTPSQTA  
 WSLGLMAAGDVHSLAVRRGVRVLLDHQNGWGTWEEKHFTGTGFPRVYLYRHYGRHFPLWALGV  
 WALGVYSRLSSGQKTRQEBERRHSSPGDLHLPWLERIGRR

>seq\_ID 71

MIKNFTALWPIRRVRKGVSVTSQDGHSSANGASKPDFEVRPHVDLETAIHRSSQSFLLKEQKPEGY  
 WVVELIVDSTLVSDTIAYHHWNGKVDMEWQRKAVNHIFSMQLPDGGWNIYGGPAEINATVKA  
 YLALKLAGVPMVDRMLRARSVALSMGVPRMNTFSKLYLALLGLFPWNYVPTIPCEVILIGKW  
 FHVNFYEMSSWSRSMVPLAIINHFKPTRKLNQVQKLDDELVEPEGYHERDLALPDPPELTPRNF  
 FLWLDKHLKFAELWVQAGIHPFRRALKKCEHWMLERFEGSNGLAAIPAMLSLIALKALGYP  
 GDHPVVKRAEKELKNLEHETADTVRIEPCFSPVWDTAIVAI CLHESGIPSDHPALKKSAEWLIDK  
 EIRFRGDWYFKNPVDVPSGWVFEFENKWNPDVDDTAMVLLALRKIPTSDVKRRDECFQGRGL  
 KWMMAFQCKDGGWAFFDKCTKGI LKLEKVPFADHNAMLDPECADITARILELLGYEGVGDHP  
 QIKKALQFIQEBQEDDGSWYGRWGVNYIYGTWQVLRGLRALNINMNQPWLLKARDWLESVQH

## Enzyme Sequences

EDGGWGERCNTYDDPVFKGGPSTASQTAWAVMGLCTFDDPQRPSLMRGIDYLIKTQNSDG  
 SWTEHEITGTGFPVFLYKYDMYRNSWPLLALATYRNLVYASSEKTANGHTNGHSVQLPEALKT  
 PPAPK

>seq\_ID 126

MNKKSAMKLLKKAKNHVSVLLQPTDALNRVMKRRFSLQSPEGYVWFALEADVITIPSEYIMFNR  
 FLGRKMDKGLAERLGNIRAKQADGGWPLHDNDGPVNI SASVKAYMALKMLGDNKDAEHM  
 VRARQILAKGGAETANVFTRI CLATFGQIPWHCPPAMP I EIVLLPKWFFPHLDKVSYSRVSVI YP  
 LLIIYAKQPVCRLRPEEAVPELFCKPABEHIHIDKYRDKGWRKLNFI LLDRVLKRTIHLVPKSINKK  
 ALNYAEKWTRHEMAGRGGI GAIFPAMANAVMALSLLGYDESDDPARGMQSVDDLMTVDKPHV  
 PEKSPWEHTVITGGAELSAAP ELDI SPDHGTAENLEQAMCQPCNSPIWDTCLTLSAMMEAGEN  
 QDSKSTQQALNWLWQQIFFRGDWISKAPKLEGGGWAFFQENTFYPDLDDTAMVLMAMCRA  
 GVLDQPEHRENF I KGVNWL I GMQSSNGGWAFFDIDNCAEYLNDIPFADHGALLDPPTSDLTAR  
 VIELLGLVLYDKSFRPI K DGI EFLKKEQEDDGSWFGRWGVNYI YGTWSVLCGLRQAGEDMNS  
 YVCKAVEWFENHQNKGGWGESCLS YNDKNYAGLGDSTASQTAWALLGLMAAGRVHSKAV  
 SRGVRYLLDTQKDDGSWDESLEFTGTGFPVRYLRYHGY SQYFPMMALGVYQRFSADEDTKQI  
 MMRKSPDLLGRKW

>seq\_ID 114

MIPTDPTPGSTQNRDLVAIRRAQQNLLRLQHNEGYWCGELFVDSTLCSYDYLVMHWADEIDPV  
 MEEKCVAHIRRRQLEDGGWNI YEGGPSDVNATVKAYFALKLAGHAPTQFWMQEARACTLRLG  
 GIPKNTYAKLYLALLGQFPWRYLPTVPEIMFMPRWFDFDIYEVSSSRAMLMLPLAI LNHYKV  
 TKHLPADKQLHLELYI GSEESDLGLGMQKPRFSWPNFFLFCDRLI KIMHSLPWKPKRAALAR  
 AEAWMTQRMGEGSDGLAAI FPAMLSMIALRTRYSREHPLYVKAKNDFAGLFVDDPQDFRIQ  
 PCLS PVWDTAINLVALLESGLDPHDPKIEAAVNWLKEKEVRLNGDWYVKNHHVPPSGWAFEFN  
 NVYYPD TDDTMMVLAALARAGAHESAPVETKAMFERALKWLLSFQCRDGGWAFFDKDVTQ  
 GWLEDPVADHNAI LDPTCSDLTGRVLELLGLIDYDRNCTPVRRALKFLRDTQEDDGSWYGRW  
 GVNYI YGTWQVLRGRS I GEDMRQQWI VRARDWLESQCNEDGGWGETCAS YDDPTLKGKGP  
 STASQTAWALMGLIAAADPTEPGA FDRKSI RQGVLYLSTQVADGSWVEPEVTGTGFPVRYL  
 RYDMYRNNFPLMALATYRKAREGKLPVRQRE

>seq\_ID 194

MKKA TRSVFSLDGGKI SDGSRGDSRHAGSRLLDSVTKSAAALLASRQNPDGHVFDLEADV  
 TIPAEYVMRCF IGEPLSDMASRLSAYLLERQLPDGGWPLYAVDGNANISATVKAYFALKLLG  
 HDKYAPHMVSARRMLAQQGAERSNVFTRI TLALFGQVPWHHTPAMP I EIMLLPKWFFHLSKV  
 AYWSRTVIVP LLLI LYNKQPVCLRGYSEGIAELFSTSPDMLVHLDHFRYRWRKNAFIVLDRLLKR  
 TMHLVPGRI KRRALBEAERWTRERMKGDGSI GAI YPAMANAVMALKLTCGSDSDPYLRGLR  
 AIDRLL IHGKPRAGALPADGAGTLFPVLDGASAAVDLYPASLSDTAKSHAFSCFCPCNSPVWD  
 TALS LTALSEAGGGYSPERAMEWLFNRQI ATQGDWTERCPGLECGGWAFFQYENALYPDVD  
 DTAKVMSLFRAGALERGEYPEKIAKAVRWVLMGQAGDGGWGAFFVDNHNHLYLNDIPFADHG  
 ALLDPS TADLTGRCI EMLGMLGHGPDYPPITRGI EFLREEQEPFGGWFGRWGVNYI YGTWSVL  
 SGLSQAGEDMGRPYRKAWEVLVSCQNDGGWGETCAS YDDPSLAGSGASTASQTAWALL  
 GLMAAGEADHAARAGIAYLADS FADGWERHFTGTGFPVRYLRYHGYSLFFPWALGVYA  
 RHREGKTVQEQVREGRVNGVDFVVMGGSA

>seq\_ID 154

MMANATDTIELPSPRAADRIVPMTD IDQAVDAHAALGRRQDDGHVFELEADATIPAEYVLL  
 EHYLDRIDPALEERIGVYLRRIQGDHGGWPLYHGGKFDVVSATVKAYFALKAI GDDIDAPHMARA  
 RAAI LDHGGAEERSNVFTRFQLALFGEVPWHATPVMPVELMLLPRKALFSVVMMSYVVSRTVIAP  
 LLVLAALRPRAINPRD VHVPELFTVPPDQVRDWIRGYPYRQLGRLEFKYVDIALRPAERLIPDATR  
 QRAI KAAVDPIEPRLNGEDGLGAI YPAMANTVMYRALGVPSDPRATAWEAVRRLVLELDG  
 BAYCQPCVSP IWDTGLAGHAMI EAASGPEGIRPEDTKKLLAAA EWLRRERQILNVKGDWAINC  
 PDVPPGWAFFQYVNDYYPD VDDTAVVGMMLHREGEDEPANDAEALERARQWI IGMQSSNGG  
 WGAFFDIDNNDLFLNHI PFADHGALLDPPTADV TARCISFLAQLGHPEDRPVI ERGIA YLR TDQEREG  
 CWFGRWGTNYI YGTWSVLCAYNAAGVAHDDPSVVRVDWLRVSVQREDGGWGEDCAS YEGA  
 TPGIYTESLPSQTAWAVLGLMAVGLRDDPAVMRGMAYLTRTQKDDGEWDEEYPNAVGFPKVF  
 YLRYHGYRQFPPLLALSRYRNLASSNSRHVAFGF

>seq\_ID 156

MLIYSDILEKEDRVSETLSRQSVPEDEINHAI EGAQAALGGKQKSDGHVWVFELEADATIPAEYVLL  
 LEHYLDRIDPEKQAKI GYVYLRRIQGHGGWPLYHGGKFDVVSATVKAYFALKAI GDDINAPHMRIA  
 REAILDHGGAARTNVFTRIQLALFGEVPWDATPVMPVELMLLPRKAFPSVVMMSYVVSRAVIAP  
 LLVLMALRPKAINPRGIVHQLFVKPPSEVKDWIRGYPYRQVGRFFKHLDSALRPVLP LIPRSVH  
 KKALKAASDFIEPRLSRGGLGAI YPAMANTVMYRAQGVPSDPRAKTAWDAIQDLDVHDGHE  
 IYQPCVSPVWDTGLSGLAMI EAASGPGTKTKE TLAAALKKSAEWLREHQI LDVKGDWAINAPD  
 LRPGGWAFQYENDYYPD VDDTAVVAMLHHRVDPENSREAI SRAREWI IGMQSTNGGWAFFDI  
 DNDHELNLNHI PFDHGALLDPPTADVSARCISFLAQLGDPDRPVILKAI EYLRSEQEPGECWF  
 GRWGTNYI YGTWSVLCALNIAGVPHDDPMVLRVAVNWLSEVQRPDGGWGEDCATYEGGTAGT  
 YKSLPSQTAWAVLALMAVGRRESEAVKRGVAYLVSVQNEKGEWQEEAYNAVGFPKVFYLR  
 YHGYKQFPPLTALARYRNLGVSNKGVYEGF

>seq\_ID 74

MEGASPTASNRI SQYAVDLRAKARA AVASTCDWLLSHQHADGHWCABLEGDSILQSEYILLLA  
 WLKGERTEIARRCAHLLKQEPNGAWTQFPGAP IDVGSSVKAYFALKL TGHDAADYVRA  
 RNAI LEAGGADKVNSTFRFLALLGQIPFELCPAVPEMVLNPNWSPINIYRISWVSRITFVPLSIV  
 WAHRAARDIVEDVSTHELFIRKPEDWELRCPGLEKPAFLSWDRFPRTADSGLKLEKYLRP  
 LRKRALRQAQQWMLDRFQQSDGPGAIFPPIVWSAIALRTLGYAEDSPETIQYCLDHLERLVLEDD

-continued

Enzyme Sequences

ETTKLQPKSPVWDTSI TLRALAAAGLGLAQEPTCRGVWLLSKEVRVPGDWTNNVDCEPGG  
WFFEYENAFYDNDTSMGIMALADQLAAANI TLEVHPGETLANTSVVVGGRGIAEQLAGSSA  
AMMEQAAAATRRVAWMMAMQNKDGGWGFADKNNDAEFLCHVPPADHNAMIDPSTPDLA  
RVIESFGRLGVTIESPGKLDGTVRRAVAYIRANQLSDGWSFGRWGVNVI YGTWQCLVGLRAVG  
VPANDPAI EQGKLLWLLAHQQA CGWGESCEYEDPSLRGQGSPTASQTAWALLGI IAAGGAN  
LAEVHVGVQYLMDTQREDGAWDEIEFTGTGFPFRVYLYKHYHYPI YFPLLLALAEWNRATARS

>seq\_ID 326  
MFDTISFDALDQAI SRAHARLSAEQRADGHVVELEADATI PAEYVLEHFLDRIDPELEARIG  
VFLRGI QGNSPQNGGWP LFDGAMDI SASVKAYFALKAI GDDPDAPHMRRAREAI LARGGAA  
RTNVFTRI QALALFGAVPWRACPVMPVEIMLLPDWFPIT IWKISYNSRTVIAPLLVLLTERPIARNP  
RNVRIDELFVTPPDQVTDYIRGPIRSNWGYLFKAIDSALRPLERHFPARSRKRAIQAAIDFI TPRL  
NGEDGLGAI YPAMANTVMYHTLGYSPDHPDYATAWASVRKLVTDAS YRFEGASVYQPCLSF  
VWDTSLAAHALAEAGSPGDAQLAAACDWLI PRQILDVKGDWAYRKPDPAPGGWAFQYNNAH  
YPDVDDTAVVGMILDRNGDPAHREAVRARQWILGMQSRSGGWGAFDSDNEFHLYLNHI PFAD  
HGALLDPPADVTARCI SFLAQLGHAEDRPAI ERGVAYLRREQEODGWSFGRWGTNYI YGTW  
SSLCALNAAGVAQDDPMVMVRAVEWLLARQRPDGGWGEDCETYAHAKPGYHESLPSQTAW  
ALLGLMAAGQAEEHVAARGIAWLQSVQEDDGSWTEQPYNAVGFPRVFLRYHYGYPFRFPFLA  
MARYRNLARGNSRQVQFGF

>seq\_ID 192  
MDKI KMKINQPKFRVFRGGQKAATPCPGTNERGALDRGRLSASLKHSREWLLSLQADAG  
NWVFALEADTTIASFYVMLQRFLGRPLAPELQQRRLANYLLSRQLPDGGWPLYAEDGFANISST  
VKAYLALKLLGYPTHCDPLVRARQIVLALGGAECNVFTRIALALFGQI PWRTTPAMPVEIMLLP  
RWFYFHLSKI SYWARTVVVPLLI LYAKRPVCRLEPWEGIP ELPFVTPDKLGYLDVCKPKGQWRKN  
VEIWDRLTRKMRVCVPRRLHNLALRAETWTREHMQGAGGIGAI FPAMANAVMALRTLGCSS  
PDDADYQRGLKALDDLLIDRCVPPREDTPVSPCWTGTS AAPMLDPSAPAGSHAQGGDQGIC  
QPCASP IWDTLGAL TALLEGLDARHPAVDRAVRWLLDQVVDKGDWAQRVFNLEAGGWAF  
QFENALYDLDLDDTSKVLMSLIRAGAMNPGYRQELSRAINWVIGMNSDGGWGAFFVDNNYL  
YLNDIPFADHGALLDPSADVTRGCI EMLAMAGFGRDFLP IARGVDFLRREQEDFGWYGRW  
GVNYI YGTWVALSGLIHAGEDLQAPYIRQAVGWLESVQNPDDGGWGETCYSYDDPALAGRGVS  
TASQTAWALLGLMAAGEVDNLAVRRGI QYLVEEQNRAGGWERHFTGTGFPFRVFLRYHYGYS  
QYFPLWALGLYERLSSGNPSRQOMVRRAGPAGLHLPVLDRRKLRRKRKA

>seq\_ID 72  
MKSEEVTI KPAVGLKDELNAAI TRSQSFLLECEQKPEGYVWVWELMVDSTIVSDTIAYHHWNGKV  
DPEWQRKAVNHLI LMSQLPEGGWNII YQNGPPEVNATI KAYLALKLAGIPI TDPMLKARQVALTL  
GGVPRMNTFSLYLALLGLWPKYVPTIPCEVLLLGKWFHVNIDWMSNWSRAMIVPLAI INHYK  
PTRPVKVDLSLELFGFHERDLALPKDPQSFTWRNFFLGLDQLHKFAELWVNAGIHPFRRLALK  
KCEQWMLERFEGSDGLAAI FPAMLSLIALKSLGYPDDHPVLRARERLEKLEHETKDTVRIEP  
CLSPGWDTAIAAMCLESVGPVPAEHPRLKKAAGDWLVNREVRFKADWHHKNPVDVPSGWVQ  
FNNKWNPDLDLDTAMVLLALRLIPTDHPRRRDEAFQRGLKWLAFQCRDGGWAAAYDKDCTKNI  
LEKVPFADHNAMLDPEDADI TARVLELLGFBGYALDHPQVQEAVEYLRHEQETDGSWYGRWG  
VNYI YGTWQTLRGLWALKMDMNQVWLLKARDWLESVQLPDGGWGERCNTYDDPVFKGGQF  
STASQTAWAVMALCTFGDPKRPSLVIRGIQYLI ENQNEGDSWTELETTGTGFPFRVYLYKYDIYR  
NTWPLLAMATYRKMMLDPKEVRVK

>seq\_ID 145  
MNKHKGTFSVIEGGKTTQARGSETCAIMDAADLEKVTSSVAASQLAGQQDDGHVFDLEADV  
TI PAEYVLMQRF IGREIDPEI SERLAAVMQERQLPDGGWPLYAVDGNVNI SASVKAYFALKLLGH  
DKNAPHMVRARQLI LSLGGAACNVFTRITLAFGQIPWHTAPAMPI EIMLLPRWFFHLNKVAY  
WSRTVI VPLLILYATQPI CRLQYNEGI TELFTTPDMLVHLDKFRHAWRKNVFI ALDRVLRKRTM  
HLVPGRI KQHALLAEERWTRARMQDGGIGAI YPAMANAVMALKTLGCSSDDADYLRGLEAVE  
DNLMVHRNLKTGTIPMDSSGGIAIDNSSAAPELSPTYLTDTAGNTEFSFCQPCNSPIWDTCMS  
LSALCESGYAENNSGVTDRAI KWLFSQQAATPGDWS EKCPGLESGGWAFQYENSRYPDVDDT  
AKVLSLFRAGALEKPEYREKIERAIRWVQGMQSTDGGWGAFFVDNDYFYLNDIPFADHGALL  
DPSTADLTGRCI EMMGMLGHGPDYPPIARGIAYLKKQEPPGGWFGWRWGVNYI YGTWVLSG  
LHQAGENMDAPYVRKAVEWLI SCQNSDGGWGETCASYDDPSLAGSGASTASQTAWALMALM  
AAGEWRHSAVRNGVRYLTESYCNWNEKQFTGTGFPFRVFLRYHYGYSLFPVWALAVYSRYI  
NGTATVQEKVREKQFRQCLMV

>seq\_ID 127  
MLPYNQDFYNEDEALKDDHCEGAGNVSNPPTLDEAI KRSQDFLLSQYPEGYVWVALEGNPT  
ITSHVTILYKLLGI EDEYPMCKMEKYLRRMQCIHGGWELFYDGGQLSVTIESYVALRLLNVPT  
DPALKKALKFI IDKGGVXKSRMFTKICLALLGCFDWRGIPSLPWNVMLLPWFLSSI YETACWA  
RGCVVPLI VVDPKPFVKSPEVSDLEYAEGREHACKTLPFCGDWTSHPPIAVDRVFKMMER  
LGVVFPQQWGI REAEKWLLEQEDTGDPLGVYPPMFYSVVMCKTLGYEVTDPVVRALLSFK  
KFSIERADECSVQSSLPVWDTALVVRSLVESGLPPDHPALQRAGEWLLQKQITKHGDWSFKN  
QSGVAGGWAFQFNRWYDLDLSDAVVVMALDCLKLPNEDEVKNGAITRCLKWIISMQCKGGG  
WAAFDKDNHQHWINSTPFSDLKAMVDPSTTDI SARVLEMVGRKLKLGTSFDEAHFLPPEIAR  
GLVYLRREQENEGCWFGWRWGVNYI YGTGALVALSLVAPMTHHEEIIARGARWLQVQNMHG  
KKINGPQDGGWGETCFSYNDPALKGQDVS TASQTAWALQGLLAAGDALGKYEVEISIGHV  
QYLLSTQRKDGSWHESQFTGGGFPIHFYLYHYFAQHFTLSSSLARYRTRLQASKIKPPIP

>seq\_ID 166  
MNTEPRFSAPETLRAIAGRALGRHQRRDGHVVELEADATI PAEYVLEHVMRDRITPERQA  
RIGAYLRRIQGHEGWPMFHAGEFNI SASVKAYCALKAIGDDPQAPHMVRARQAILGHGGAER

-continued

## Enzyme Sequences

ANVFTRIQLALFAGIPIWRGVPMVPEIMHLPKWFFFNIWAMSYWARTCVVPLLVLQARKPRAR  
 NPRQVSFDEIFRTEPDEVRDWIRGPIYRSRWGVVFKHIDTVLRWTEPLFSKVARESAIFKAVDFV  
 EERLNGEDGLGAIYPAMAYALMMYDVLGYPEDDPRCVTIWKAIKLLIETDEEVYCVPCVSPV  
 WDTLSLGHAMIEAARTGGIEAQAELEDAACDWLVARQVKDVRGDWAETRPDAEPGGWAFQYR  
 NDHYPDVDDTAVVAMLLHRNGRPEHAEAEIKARRVWVGVQSRNNGWGAFDADNDREFLNHI  
 PPSDHGALLDPPPTADVTGRCISFLSQLGHEEDRPVIERALAYLRAEQERDGSWYGRWGTNYV  
 YGTWTVLCGLNAAGIPHDDPMVRRAVDWLVSIQRADGGWGEDERSYDVGHYVENAESLPSQ  
 TAWAMLGLMSVQADHPAVLRGAAYLQRTQGPDGEWQERAYNAVGFPRVYLYKHGYYRLLF  
 PLFALSRLHNLQRGNSREVSFGF

&gt;seq\_ID 21

MSGEVVRVAGDALAEDAGRAAAAASQYLRTQQRDHWRAELESNTVTAEYVLLRQALGLDLE  
 ERRDALVRYLCSRQKADGSFGIASTLPGDVTSTAEAYLALRLLGLDREDERLRAAERFIRGAGG  
 LARVRVTRINLALFGLFPWEAVPTVPAELIFLPRWAPVNVYRLASWARSTMVPLVLFHHRPV  
 PALPGGAGSDWLDHLWLGPGDKRVPYRYSVMEVTVRRHGPGWKAFNAADAWLRVHDLRLH  
 LPPLGRLRTEALRACEEWILARQEASGDWAGIFPMLNGVLAHVAGHGLDAAPVRRGLEAIE  
 RFAVSDREGFRIEACQS PVWDTI LALIGLLDSGESPTDPRLVAARRWIEGMQLTNDWGDWVY  
 DPRGEPGGWAFYANSWYPDVDDTAAVIVGLLKHDPASRAGETVRRAAAWVASMQRNDGG  
 WAAFVNDNRLFLNEIPFSDMDSLCDPSSPDVTGRVLEAFGMLDAPHLRAACRRGVAYLRRA  
 QEPEGSWYGRGVNYVYGTSNVNLGLARQRPASDPMVARALGWLDSVQNAADGGFGEGL  
 SYADRAAMGRGPSTASQTAWGVMLLARYAADAARVRIAWLVERQLADGAEQGSWEIE  
 AFTGTGFPFRHFYLYRHYFPLMALGRFCAQGRG

&gt;seq\_ID 111

MSYEWTEPVRPGRHRAVSPVQNFQCQSLAPAIQRACDALFSQQAADGFWCGELTADTTLESYD  
 ILLQLWLNQPDHGWNPPTPRIDRAGRSILERQLPDGGFNIYAGGPSEVSATIKAYCALKLAG  
 LDPHSPPLRRARERILALGGQAANSYVKINLSLFLYPRKHVPSVPEIIVMLPGNVLYEMS  
 TRSILVPLSIVQARGSNRRAPNGFNLELLLPVGLALPKRKGALVLFHHLDRMFKVWEKRGSE  
 RIRGAAIREAERWL IARTHYTEGLGAIYPAMMYFIMALDALGYAEDHPDRSEAIRHFESLLIETDD  
 RFLFQPCVSPVWDTAICAFALGEAGNTDDPRMTLADWLI SKEVRRKGDWSIKRPTDEPSGW  
 APEFANEFPYDIDDAMVLLALMHANGSNPEAQAARAVNWLAMQSSDGGWAAFVDN  
 NWAMLNQVFPADHNAMLDP TPCPDITGRVLECLCRRGMAGHDAARRGVAYLLQAQEKDGSWY  
 GRWGVNYIYGSFLAMRGLTTS GAPGSQDAVDRAARWLRAIQNPDGGWGESCSASYARDGYVA  
 APSSASQTAWALLGLCAAGDRDSAQFRRGVEYLLTLQAPDGKWPEGATGTGFPNVFYLYT  
 MYRDYFPLLALSQV

&gt;seq\_ID 157

MPKDI PADLASEAI SGMLEQAVLRASMALHRKQQT DGHVVELEADATIPAEYVLEHFLDRI  
 DDDLERKIGVYLRRIQGDHGGWPLFHEGAFNLSASVKAYYALKAI GDDPDAPHMRRAEALIAA  
 GGAERSNVFTRIQLALFQGIPIWRGVPMVPAELMIAPKWFPIIMWIKVSYWRSVTVIAPLLVLM  
 KPKARNPRNVHRELFLHDPDRIDWIRGPIFRSGWGHFFKYLDVSLRVVPEVALKPMRPRISR  
 LAVDFVRERLNGEDGLGAIYPAMANSVMYDVLGYSPDHPEAAIAWESVRLKLVKEDEAYCQ  
 PCLSPIDWDTGLSGHMAEAEGAVSPGVAAACDWLRNRQITDVVGDWAEIRPGVQPGGWAFQ  
 YNNAHYPDVDDTAVVAMLLHRQGDPAHEESIRKAREWIIGLQCRDGGWGFADNDKDYLNH  
 I PFADHGALLDPPPTADVTARCISFLAQLGNPEDKPVIDRAMAWLRKEQEDGSGWFRWGTNYI  
 YGTWVSLCAMNVAGMPHDDPAIRRAVNFVATQREDGGWGEDEEYD PASGAQPGRYKEST  
 PSQTAWALIGLMAAGEAEHEATRRIAYLQATQKPDGEWDEAAAYTAVGFPRVYLYKHGYYR  
 FFPLMALSRYNLRSNNMKVVSFGF

&gt;seq\_ID 205

MNQAATITRPQDETLLTSARRPAQPALPDPLDAGIAHVVESLLAQQQSDGHVVELEADATIPAE  
 EYILMVHYLGETPDLVLEGGIANYLRRIQNADGGWPLFHAGASDISASVKGYFALKMAGDNPEA  
 EHMRRARAAMHAMGAEASNVFTRTLALYGVMPWQAVPMPVPEIMLLPEWFPFHLKSVSYW  
 ARTVIVPLLVNLSLRPQARNPKIGIDELFVRPCQATRLPRRAPHQSPLWVGVFRTLDVVRMA  
 EPLFPRGLRQRAIERAREFTVERLNGEDGLGAIYPAMVNSVLMFVLDVLPVPESDPNRAIARRS  
 IDKLLVTKDDEAYCQPCLSPVWDTSLAAHALLEVGEPRITAAAARGLDWLLPLQELELRGDWTVRR  
 PNVRPGGWAFQYANPHYPDVDDTAVVAAMDRVDKDRSNRYDEAVSRACEWIVGMQSSN  
 GGWGAFAPEPENTHLYLNNIPFADHGALLDPPPTADVSARCLAMLCQLGQMPANSEPAARALRYLL  
 DEQEADGSGWFGWGTNYIYGTWSALCGLNAAGIGTDAPEMKRAAQWLLSIQNEGGWGESG  
 DSYKLEYRGYKAPSTASQTAWAMLGLMAAGADHPALVRGVEYLLRQTQASHGFWDPEYFT  
 AVGFPRVYLYYHGYSRFFPLWALARFRNLRLDGNRAISWGL

&gt;seq\_ID 218

MKTDGNTTLDTTISMEELERTVKSAYEALAKDQDDGHWIELEADVTIPAQFILLEHTLKDIDE  
 ELEQKIANYLRRQCSREHWGWPVYGGEFNISASVQAYFALKMTGEDINAPHMVRAREALIAH  
 GGPEYANVFTRIQLSLFGEASWLATPFMPVPEIMLLPRWYFYSIWNMSYWSRTTVAPLLIVADLK  
 PKAINPRNVHIEPELPTPPDKVKTWIGHFPFRSKWGHVFKFIDTAIRPFTRFVPSFLHKKAYKAAL  
 DFIEPRLNGVDGLGAIYPPMSYSAVMYRALGIPDDDPRAATNWEALKGLLVIKEREAYCQACVS  
 PVWDTALSGHALMEASFGPDGINADRTEKIDRAAHWLRRAHQVLLNVVGDWAINNPNLQPGGW  
 AFQYGNDYYPDVTAVAAAMLHRRQNLPENEEALDRARKWIIIGMQSSNGGWGAFDIDNDKQI  
 LNDIPFADHGALLDPPPTADVSARCSLLAELGHPEDRPVIERGIKYLKKEQEDGSGWFRWGTN  
 YIYGAWSVLCAFNASGVPHDDPSVLKCVNFKLSVQREDGGWGESCEYEGSAHGVTYTESLPS  
 QTAWAVLGLMASGRRTDPAVKRGIWVLIQHQQDNGEWAEEPFNAVGFPRMPLYLHYLGYKQF  
 FPLLALARYHMEKSGTNNVSFAF

-continued

## Enzyme Sequences

>seq\_ID 11  
 MLPYNQDHHFGKVAENATMPPTLDEAIERSQDFLLSLQYPEGYWWABLEANVTTLTAQTIMLYKI  
 LGIDHKYPIHKMPTYILRTQRAHGGWEIFYGDGGCLSTTIGAYMALRILGVPKTDVPLQKALKLIH  
 SKGGVTKSRMFTKI CLALLGCYDWKGI PSLPPWLVLPSWFFPSLYDTASWVRGCVVPLTII FD  
 KKPVYKLNPLLCLELYSEGKGRVHLSFIPGDWTSNFFVGLDHFVKYEMENLGVVPPRQWGI  
 KEAERWTLERHEDSGDFHGIYPPMFYSIVSYLLGYEITDPVVHRALESMRGFTVEREDECVV  
 QSCI SPMWDTAFVIRSLAESGLQDPDHPALQKAGEWLLQKQATQGHGNYFYKRTGRAGGWAF  
 QFFNRWYPDVDDSAAVSMALNAIKLQDDVKKGAIKRCAEWISVMQCKDGGWAAAYDCNDR  
 EWLNCTPFGDLKAMIDPNTVDVTARVLEMPWRVGRVKEAGDASAILPPRAIARGLAYLRREQETEG  
 CWYGRWGNYYIGTSGALMALALVAPSTHKEEIERGARWLVEVQNKRGTKGANGYSHTNGA  
 REGGVAMNGCNKMGAPEDGGWGETCFSYNDITLKGGRNEVSTVSQTAWALQGLLAAGDALG  
 KYEVESIEHGVQYLLSTQRKDGSWCEKHFTGGGPRFFYIRYHLYAGHPLSALARYRDRVRA  
 GKMAK

>seq\_ID 214  
 MDATAPLRDPGAPS AENCSDVRRRELDVIGESCRWLGERQNDQGHVFELEADATI PAEYILL  
 NHFLDEIDDAREAR IASYLRAIQGKHGGWPLFHDGDFDMSATVKAYYALKLTDGVDVDEPHMVR  
 ARQATLEHGGGAERTNVFTRFTLAMFDQVPRACPVTPVEALLPRFAPFHWSKVSYSWRTVM  
 TPLMILYSRRARAVNPRGIGVRELFRRDPEVIRDWLKNPTGHWIGDALIQIDKVLRLVIEPAIHWAF  
 RDRAEKWALDFIEERLNGRDGLGGIYPAIANTLMAYHTLGYAKDHPGYRIAREAVDGLCTPHAK  
 GEYVQPCLSPVWDTCLASHAIQEAGQSAGDRAVDQSNAWLRERQVLDVVGDWKSNGRHLRP  
 GGWAFQYNNPHYPDVDDTAVVVMALARSKEDANREAIARAEWII GMQSSNGWGAFDAE  
 NEHDFLNHVFPADHGALLDPPVTDVSARCLGMLAQLGRPKTDPVVARGLDYLRWREQADGS  
 WFGRWGTYNYIGTWSALNAFNAVEWDMTDPRI CKAVDWLKSQRDDGGWGEDCATYWKER  
 RSVSKASTPSQTAWAVLGLMAAGEVDSPEVERGIRYLLLEAPRDGGKWEELLYNAVGFPRIFYL  
 RYHGYSAYFPLWALARYRNLTSGNCKRTIHGM

>seq\_ID 73  
 MPEEAILTETHPLDATTIETAITRARKALLGEQRADGHFVFELEADVSIPEYILFYHFIGRPAPAE  
 LEAKIGHYLRARQSAEHGHWPLFDGAFNIISSSVKAYFALKAI GDTDPMPHMQRARTAILAHG  
 GAAANVPTRSLALLFLGLI PWHGIPVMPPIEMHLEPEWPFPHIAKISYWGRTVLVPMVHVHALKPK  
 PANTCTIRIDELFVIPPQVHRHWGSPGKRFPWTAIFAGIDKVLQIAEPPYFRRSRQSAIDKAVA  
 FVTKRLNGEDGLGAIYPAMAYSALMYLSIGRSLSDPHQLVLKAIIDKLVVVKDHEAYVQPCVSPV  
 WDTALASHALMEAGDGDKPI LDSLKKGGLAWLKLQVTDIAGDWANKKPDVVKPGWAFQYGN  
 AYYPDLDTTAVVVMAMDRARDRWEI DEEDNFRPSIARAREWIVGLQSENGGFGAFDADNDRD  
 YLNAIPFADHGALLDPPVTDV TARCISMLTQLGKPENSETLRRRAIAYLFAEQEKDGSWFGRWG  
 LNYIYGTWSVLCSLNAAGI AHDAPEVRRRAVAVLRTIQNEDGGWGEDAESYALDYAGYQQAPS  
 TSSQTAWAVLGLMAAGEKDDPAVARGIAYLTRTQGEDGFWTEKRFATGFPFRVFLRYHGYS  
 KFFPLWAMARYRNHLHNGNHASVLTGM

>seq\_ID 103  
 MNDMTEMHTLDATAVPAAPAAADAPAPSAATTGLDAAVARATDALLAAQNADGHVVELEAD  
 STIPAERYLLVHYLGEENAELEQKIARYLRRIQQPDGGWPLFTDGAPNISASVKAYFALKVIGD  
 DENAEHMQRARRAIHAMGGAEMSNVFTRIQLALYGVVWPYAVPMPVVEIILLPQWFPFHLSKV  
 SYWARTVIVPLLVNNAKRPVAKNPRGVRIDELFKSAPVNTGLLPKQPHQAGWFAFRAVDGV  
 LRLADGLFPRYTRERAI RQAAAFVDERLNGEDGLGAIYPAMANAVMMYAALGYPEDHPNRAI  
 RQSI EKLLVVEEYAYCQPCLSPVWDTSLAAHALLETGDERAREAAVRGLDWLVPRI LDVRG  
 DWISRRPHVRPGGWAFQYANAHYPDVDDTAVVVMAMDRVAKHDQTDAYRESIARAREWVVG  
 MQSSDGGWGFEPENTQYLLNIPFSDHGALLDPPVTDVSGRCLSMALQGETNASSEPARAFD  
 ALDYMLKEQEPDGSWYGRWGMNYIYGTWTALCSLNAAGLGHDDPRVKRAAQWLLSIQNPDG  
 GWGEDGDSYKLDYRGERAPSTSSQTAWALLGLMAAGEVDNPAVARGTGHLLGTQREHGLW  
 DETRFTATGFPFRVFLRYHYGRKFFPLWALARYRNLRKAGAAARVTVGM

>seq\_ID 95  
 MNDMTEMHTLDAAAAPAADAPAVTAVTAGLDAAVARATDALLAAQNADGHVVELEADSTIPA  
 EYVLLVHYLGEENAELEQKIARYLRRIQQPDGGWPLFTDGAPNISASVKAYFALKVIGDDENA  
 EHMQRARRAIHAMGGAETSNVFTRIQLALYGVVWPYAVPMPVVEIILLPQWFPFHLSKVSYW  
 ARTVIVPLLVNNAKRPVAKNPRGVRIDELFKSAPVNTGLLPKQPHQSTGWFAFRAVDGVLRV  
 DGLFPRYTRERAI RQAVAFVDERLNGEDGLGAIYPAMANAVMMYAALGYPEDHPNRAIARQSI  
 EKLLVVEEYAYCQPCLSPVWDTSLAAHALLETGDERARDAVRGLDWLIPRQILDVRGDWIS  
 RRPVHRPGGWAFQYANPHYPDVDDTAVVVMAMDRVAKLDQSDAYREQIARAREWVVMQSS  
 SDGGWGFEPENTQYLLNIPFSDHGALLDPPVTDVSGRCLSMALQGETNASSEPARAFD  
 YMLKEQEPDGSWYGRWGMNYIYGTWTALCSLNAAGLGHDDPRVKRAAQWLLSIQNPDG  
 GEDGESYKLDYRGERAPSTSSQTAWALLGLMAAGEVDNPAVARGIYLLGAQCEHGLWDET  
 RFTATGFPFRVFLRYHYGRKFFPLWALARYRNLRKRANTRTVTVGM

>seq\_ID 106  
 MNDLTDMPTLAADSAAADLDAVARATDALLAAQADGHVVELEADSTIPAERYLLVHYLGET  
 PNLELEQKIGRYLRRIQQPDGGWPLFTDGAPNISASVKAYFALKVIGDDENAEMQRARRAIHA  
 MGAEMSNVFTRIQLALYGAIPWRVAVPMPVVEIILLPQWFPFHLSKVSYWARTVIVPLLVNNAK  
 RPIAKNPRGVRIDELFIDPPVNAAGLPRQGHQSGWFAFRRVVDHALRAVDGLFSPYTRERAI  
 QAVAFVDERLNGEDGLGAIYPAMANAVMMYDALGYPEDHPNRAIARRSVEKLLVHDDDEAYC  
 QPCLSPVWDTSLAAHALLETGDPRAEDAVVRGLEWLRPLQILDVRGDWISRRPNVRPGGWAF  
 QYANPHYPDVDDTAVVVMAMDRVAKLRHSDAYREAI SRAREWVVMQSSDGGWGFEPEN  
 TQYLLNIPFSDHGALLDPPVTDVSGRCLSMALQGETAANSEARRSLDYMLKEQEPDGSW  
 YGRWGMNYVYGTWTALCSLNAAGLGPDDPRVKRGAQWLLSVQNKDGGWGEDGDSYKLDY

## Enzyme Sequences

RGYEQAPSTSSQTAWALLGLMAAGEVNHFAVARGIDYLIAEQKEHGLWDETRFTATGFPRVFY  
LRYHGVRKFFPLWALARYRNLKRANATRVTVGM

>seq\_ID 87

MNDLTEMATLSAGAVPAGVDAVARATDALLAAQADGHWVYELEADSTIPAEYVLLVHVLGE  
TPNLELEQKIGKYLRRIQADGGWPLFTDGAPNISASVKAYFALKVIGDDENAEHRARRAIH  
AMGGAEMSNVTRIQALALYGAIPWRVPMMPVEIMLLPQWFFPHLSKVS YWARTVI VPLLVLNA  
KRPLAKNPRGVRIDELFDPPVNAAGLLPRQGHQSPGWFAFFRVVDHALRAVDGLFSPYTRERAI  
RQAVSFVDERLNGEDGLGAIYPAMANSVMYAAALGYAEDHPNRAIARKSVEKLLVVHDEAYC  
QPCLSPVWDTSLAAHALLETGDARAQEAVALRGLWLRPLQLLDVVRGDWISRRPNVRPGGWA  
FYANAHYPDVDDTAVVVMAMDRQAQKLTQSDTYRESMARAREVWVGMQSSDGGWGAPEPEN  
TQYYLNNIPFSDHGALLDPTADVSGRCLSMQLGETPLNSEPARRALDYMLKEQEPDGSWY  
GRWGMNYVYGTWALCSLNAAGLTPDDPRMKRGAQWLLSIQNKDGGWGEDGDSYKLNRYG  
YEQAPSTASQTAWALLGLMAAGEVNHFAVARGVDYLVAQQNEGLWDETRFTATGFPRVFY  
LRYHGVRKFFPLWALARYRNLKRANATRVTVGM

>seq\_ID 107

MNDLTDMANLSAGTVPAGLDASVARATDALLAAQADGHWVYELEADSTIPAEYVLLVHVLGE  
TPNLELEQKIGRYLRRIQADGGWPLFTDGAPNVASVKAYFALKVIGDDENAEHRARRAIH  
HAMGGAEMSNVTRIQALALYGAIPWRVPMMPVEIMLLPQWFFPHLSKVS YWARTVI VPLLVLNA  
AKRPLAKNPRGVRIGELFDPPVNAAGLLPRQGHQSPGWFAFFRVVDHALRAADGLFSPYTRER  
AIRQAVSFVDERLNGEDGLGAIYPAMANAVMMYDVLGYPEDHPNRAIARKSIEKLLVVHDEAY  
CQPCLSPVWDTSLVAHALLETGDARAQEAVALRGLDWRPLQLLDVVRGDWISRRPNVRPGGWA  
FYANAHYPDVDDTAVVVMAMDRQAQKLTQSDTYRESIARAREVWVGMQSSDGGWGAPEPE  
NTQYYLNNIPFSDHGALLDPTADVSGRCLSMQLGETPLNSEPARRALDYMLKEQEPDGS  
YGRWGMNYVYGTWALCSLNAAGLTPDDPRVKKRAAQWLLSIQNKDGGWGEDGDSYKLNRY  
RGFEPAPSTASQTAWALLGLMAAGEVNHFAVERGIGYLIAQQNDEGLWDETRFTATGFPRVFY  
LRYHGVRKFFPLWALARYRNLKRANATRVTVGI

>seq\_ID 212

MESGNNKQPAAGALDASIESATNALLGYRQPDGHWVFELEADCTIPAEYVLLRHYLGEVDA  
ALEAKIANYLRRVQGAHGGWPLVHDGFPDMSASVKGYFALKMIGDDIDAPHMAKAREAIRSRG  
GAIHSNVFTRFLSMFGITTWRSVPVLPVEIMLLPMWSPPHLNKISYWARTTIVPLMVAALKPR  
AVNRLDIGLDELFLQDPKSIKMPAKAPHQSWALFKLFAIGDAVLRITIEPLPKLRDHAIKLAVDF  
VEERLNGEDGLGAIYPMANVMMYKVLGFPEDHPRAITRRGIDKLVIGEDEAYCQPCVSPV  
WDTALTCHALLEVGGAAVPPAKRGMDWLLPKQVLDLKGDWAVKPNLRPGGWAQYNNAH  
YDLDLDDTAVVVMAMDRSRRATGSRREYDAIARAREWIEGMQSDGGWAAAFVNNLEYYLNNI  
PFSDHGAMLDPTEDVTRARCVSMLSQLGETAASSKAVADGVEYLRRTQLPDGSWYGRWGLN  
YIYGTWVLCALNAAGVDHQPVIKAVTWLASVQNPDDGGWGEAESYRLNYTRYEQAPTTA  
SQTWALLGLMAAGEVDSPVVARGVYELKSTQTKGLWDEQRYTATGFPRVFYLYHGVAKF  
FPLWALARYRNLRSNNSKVVGVGM

>seq\_ID 101

MNDLTEMATLSAGAVPAGVDTAVARATDALLAAQADGHWVYELEADSTIPAEYVLLVHVLGE  
TPNLELEQKIGKYLRRIQADGGWPLFTDGAPNISASVKAYFALKVIGDDENAEHRARRAIH  
AMGGAEMSNVTRIQALALYGAIPWRVPMMPVEIMLLPQWFFPHLSKVS YWARTVI VPLLVLNA  
KRPLAKNPRGVRIDELFDPPVNAAGLLPRQGHQSPGWFAFFRVVDHALRAVDGLFSPYTRERAI  
RQAVSFVDERLNGEDGLGAIYPAMANSVMYDVLGYAEDHPNRAIARKSIEKLLVVQDEAYC  
QPCLSPVWDTSLAANALLETRDARAEDAAIRGLEWLRPLQLLDVVRGDWISRRPHVRPGGWA  
FYANAHYPDVDDTAVVVAAMERAQQLKQNDAYRDSIARAREVWVGMQSSDGGWGAPEPEN  
TQYYLNNIPFSDHGALLDPTADVSGRCLSMQLGETPLNSEPARRALDYMLKEQEPDGSWY  
GRWGMNYVYGTWALCSLNAAGLTPDDPRVKKRAQWLLSIQNKDGGWGEDGDSYKLNRYG  
FEQAPSTASQTAWALLGLMAAGEVNHFAVARGIDYLIAEQNAEGLWDETRFTATGFPRVFYLR  
YHGVRKFFPLWALARYRNLKRDNTRTVTVGL

>seq\_ID 112

MSAPSHVGNTEHAELATRKAMAYLTCQERDGHWCALDTADTTLESYILFQLWLYPPQDG  
KWPEETRPLIRKAVNSILERQLPDGGFNICVGGPSEVSASVKAYVAMKLAGLPPEDDRMARLR  
ERILALGGIQAANSYVKVNLSDLFLYPREFSPSIPPEVALLPFDLLYQMSAWTRAIVI SLGIVHAAN  
PRRPAPAGFNLOELWLPVGSPEFRDPSFFTWHNTFLTVDKALKLWERYGSKAVRRRAVEKA  
KTWMIERLHSDGLGAIYPMMYSVMALDVLGYAKDDPLRVEALRHFNNLMVDDGDRFFFPQ  
CFSPVWDTAIGAYALVQADPSHEAIPAADWLIAKEVRRKGDWSVKRPNTEPSGWAFEYSNE  
YYPDIDDTAMVMLALGETRASNTAQAAACKRGLAWLLAMQSSDGGWAAFDANNWFEFLSQ  
VPFADHNAMLDPCTADI TGRVLEALASQGLDRNHKAVRRGAEWLIRHQENDGSWYGRWVA  
YIYGTFCALRGLAASGENDREAHILRAGEWLRSIQNDGGWGESCKSYDNRIFTGGPSTPSQT  
AWAILGLIAGGDANSLSVQHGIEYLLETQRSDGSWDEQFATGTGPRVFYLYNHYMKDYFPLL  
ALASFVKARAGSNG

>seq\_ID 83

MNDLTEMATLSAGTVPAGLDAVASATDALLAAQADGHWVYELEADSTIPAEYVLLVHVLGE  
TPNLELEQKIGRYLRRVQADGGWPLFTDGAPNISASVKAYFALKVIGDDENAEHRARRAIH  
HAMGGAEMSNVTRIQALALYGAIPWRVPMMPVEIMLLPQWFFPHLSKVS YWARTVI VPLLVLNA  
NAKRP IAKNPRGVRIDELFDPPVNAAGLLPRQGHQSPGWFAFFRVVDHVLRAADGLFSPYTR  
RAIRQAVSFVDERLNGEDGLGAIYPAMANAVMMYDVLGYAEDHPNRAIARKSIEKLLVVHDEA  
YCPCLSPVWDTSLAAHALLETGDARAEEAVIRGLEWLRPLQLLDVVRGDWISRRPHVRPGGWA  
AFQYANAHYPDVDDTAVVAVAMDRVQKLNHNDTFRDSTIALAREVWVGMQSSDGGWGAPEPE  
NTQYYLNNIPFSDHGALLDPTADVSGRCLSMQLGETPLNSEPARRALDYMLKEQEPDGS

-continued

## Enzyme Sequences

WYGRWGMNIVYGTWTALCALNAAGLTPDDPRVKRGAQWLLSIQNKDGGWGEDGDSYKLN  
 RGFYEQAPSTASQTAWALLGLMAAGEVNNPAVARGVEYLIAEQKEHGLWDETRFTATGPRV  
 YLRHYGKRKFFPLWALARYRNLRKRDNATHVTFGL

&gt;seq\_ID 175

MLQTEAITTEGLRFRSLAPDDPLLRVKQALKLKSQHSREEMHSDGHWCGEVKTNATTSAEH  
 VLLCQALDINLDADREAFISWFRCTQGADGGWSTAPDQAGDISVTVEAYLALKLGLSEDDAAM  
 RSARDFAIAGGVARVRIFTRIYLAMFGLFPWAAVPELPELILLPSRVPVSIYHWSAWARATVV  
 PLLIISHHRPIYALPGGKATCSYLDLWCDPRNKMVYNHDKPTAWRSDPFALIFTLADSI LHR  
 LDGLRSFNPLRRFALRKCVDWILEHQEDMGDI GDIMPLHGMALALRLEGYPLHSDPIHRGLEA  
 IERFAYRDQQKRIQTTVSAFWDTSLMLVALGDAGMASPWLTRSLGWLQQHQLGNVGDW  
 KVNNPGLKAGGFSFGYFNTWYPDVDDTASAVLAIIRQDERLVCSASVLDALNWLGMQNTDG  
 GWGAFDRDNMLKFLNKIPFSMEAFCDPSTPDVTGHVLEAFGI FLAVSARQQSPTKADVLTDR  
 VSARRAI CYLSDTHVSSGGWYGRWGCNYI YGTS AVL CALAYFGSKSDTLSGVR SVKDAVNQ  
 AIRWLETVQNQDGGWGETVNSYKDP SRAGSGPSTASQTAWAIMALLPYPSTEV IQRGVEYL  
 LRTQTKTASQGATWHEKAYTGTGFPKYFYMGSFYCHYFPMALGRYAYPCPEWHENWRPK  
 KE

&gt;seq\_ID 88

MNDLTDMATLSAGAAPAADLDAVARATDALLAAQNADGHVVELEADSTIPAEYVLLVHYLG  
 ETPNLELERKIGRYLRRIQQADGGWPLFTDGAPNVSASVKAYFALKVIGDDENAHEMQRARRAI  
 HAMGGAEMSNVFTRIQLALYGAIPWRVAVPMMPEIMLLPQWFPFHLSKVSYWARTVIVPLLV  
 NAKRPLAKNPRGVRIDELFIDPPVNAAGLLPRQGHQSAGWFAFFRVVDHVLRAVDGLFPKYTR  
 RAIQAVSFVDERLNGEDGLGAIYPAMANAVMMYDVLGYAEDHPNRAIARKSIEKLLVVDHDEA  
 YCQCLSPVWDTSLAAHALLETGDPRAEDAALRGLWLRPLQILDVGRDWISRRPNVPRGGW  
 AFQYANAHYPDVDDTAVVAMAMDRVAKLDRDAYRESIARAREWVVMQSSDGGWGAFFEP  
 ENTQYYLNNIPFSDHGALLDPPPTADVSGRCLSMLSQLGESALTSEPARRALDYMLKEQEPDGS  
 WYGRWGMNIVYGTWTALCALNAAGLTPDDPRVKRAAQWLLSIQNKDGGWGEDGDSYKLN  
 RGFYEQAPSTASQTAWALLGLMAAGEVNNPAVARGIDYLLAEQKEHGLWDEVRFTATGPRV  
 YLRHYGKRKFFPLWALARYRNLRKANATRVTVGM

&gt;seq\_ID 92

MNDMTEMHTLDATEAAPAGLDAVARATDALLAAQADGHVVELEADSTIPAEYVLLVHYLGE  
 APNVELEQKIARYLRRIQQPDGGWPLFTDGAPNISASVKAYFALKVIGDDENAHEMQRARRAI  
 AMGGAEMSNVFTRIQLALYGVVWPVAVPMMPEIMLLPQWFPFHLSKVSYWARTVIVPLLV  
 AKRPAKRNPRGVRIDELFKGAPVSTGLLPKQPHQSAGWFAFFRAVDGVLRLVDGLFPYTRER  
 AIRQAVAFVDERLNGEDGLGAIYPAMANAVMMYALGYPEDHPNRAIARRSIEKLLVVGQEA  
 YCQCLSPVWDTSLAAHALLETGDERAREAAVRGLDVLVPRQILDVGRDWISRRPHVPRGGW  
 FQYANAHYPDVDDTAVVAMAMDRVAKLDRDAYRESIARAREWVVMQSSDGGWGAFFEPEN  
 TQYYLNNIPFSDHGALLDPPPTADVSGRCLSMLAQFGETSASSEPARRALDYMLKEQEPDGS  
 WYGRWGMNIVYGTWTALCALNAAGLGHDDPRVKRAAQWLLSIQNKDGGWGEDGDSYKLDYR  
 GYERAPSTSSQTAWALLGLMAAGEVDNPAVARGVDYLLGTQREHGLWDETRFTATGPRV  
 YLRHYGKRKFFPLWALARYRNLRKANAMRVTVGM

&gt;seq\_ID 206

MTRKTI PASELDAIVRARDALLDRQHPDGHWCFELECDATITAEYIIMMHFVDEIDTALQARM  
 AKYLRAVQRLDGHWGAWDLIFGGDLDISCSVKAYFALKAAGDPPDAPHMVRAREAILARGGAA  
 KSNVFTRI LLATFGEIPWRGTPFMPVEFVLPFRWAPIHMDKVAVVARTTMVPLLVLC SIRAAK  
 NPLGVHVGELFVTPPELREYFPRKRLGQAFLVADRVRHLEPLIPALRRRAIQRVAVESEA  
 RNMGEDGFGGIFPPMVS YEMMVLLDY PEDHPLRVECKAALKLVVHRDGGSSY CQCLSPV  
 WDTAWSVMAL EQAPSDARTETAIARAYDWLTDQVLDLRGDWENNAAPSTPPGGWAFQYEN  
 PYPYDIDDSAVVLAAMLHARGKRTGQPGRYEMPVARCLDWIIGLQSRNGGFGAFDANCDRDFL  
 NAIPFADHGALLDPPTEDEYSGRVLALGITERPQDATA RERICIYLRDTPQPDGSSWGWG  
 NYIYGTWSVLAGLGLAGVDRKLPVNRNGLQWLRGKQNA DGGWGETNDSYARPELAGKHE  
 GSMAEQTAWAMLGQMAVGEGDADSVHRGAAVLLDAQNEDGFMMHPYHNAPGPPRIFHLKY  
 HGTYAYFPLWALGRYRRLAARASAMQTAKAESAESMTAH

&gt;seq\_ID 96

MNDLSMTQTLGEVLPQTLIDDHAPVAAALATGAAPVDALDAVTRATEAILAVQKDDGHVVE  
 LEADATIPAEYVLLVHFLGETPNLELEQKIARYLRRIQLPNGGWPLFTDGAMDVSASVKAYFALK  
 MIGDPEDAAMHVRARECILANGGAEANVFTRI LLALFGVVTWYAVPMMPEIMLLPKWFPFH  
 SKVSYWARTVIVPLLVNNAKRFPVARNPRGVRIDELFRGAPVTTGLLPRSGHQS KSWFAFFRAV  
 DGVLRVTDGLFPKASRERAIKA AVSFVDERLNGVDGLGAIYPAMANSVMYDVLGYPADHPNR  
 AIARRESIEKLLVVDHDEAYCQCLSPVWDTSLAAHALLETGDERAREAAERGLAWLRPLQILDV  
 RGDWISRRPDVPRGGWAFQYNNAHYPDVDDTAVVAMAMHRSAAVNTSNVNDANAIARAREWV  
 VGMQSSDGGWGAFFEPENTQYYLNNIPFSDHGALLDPPPTADVSGRCLSMLAQLGEMPATSEPA  
 RRAYDYLLKEQEDDGSWYGRWGMNIVYGTWTALCALNAAGISLEEDARIKRAAQWLVSIQNA  
 DGGWGEDGTSYKLDYRGYKAPSIPTASQTAWALLGLMAAGYVDHPAVARGIDYLRQREQRDHL  
 WDEERFSATGPRVYLRHYGKRKFFPLWALARYRNLRKRTGKRVTVGM

&gt;seq\_ID 104

MNDMTEMHTLDATEAAPAAPTATGLDAVARATDALLAAQNADGHVVELEADSTIPAEYVLL  
 VHYLGEAPNVELEKRIARYLRRIQLPDGGWPLFTDGAPNISASVKAYFALKVIGDDENAHEMQR  
 ARRAIHAMGGAEMSNVFTRIQLALYGVVWPVAVPMMPEIMLLPQWFPFHLSKVSYWARTVIV  
 PLLLVNNAKRPAKRNPRGVRIDELFKSAPVNTGLLPKQPHQSAGWFAFFRAVDGVLRLTDGLFP  
 RYTRERAIQAVAFVDERLNGEDGLGAIYPAMANAVMMYALGYPEDHPNRAIARQSIKLLV  
 GEDEAYCQCLSPVWDTSLAAHALLETGDERAREAAVRGLDVLVPRQILDVGRDWISRRPHV

-continued

Enzyme Sequences

RPGGWAFQYANAHYPDVDDTAVVAMAMDRVAKLDRDAYRESIARAREWVVMQSSDGGW
GAFEPENTQYYLNNIPFSDHGALLDPTADVSGRCLSMQAQFGETSASSEPARRALDYMLKEQ
EPDGSWYGRWGMNYIYGTWTALCSLNAAGLGHDDPRVKRAQWLLSIQNPDGGWGEDGDS
YKLDYRGERAPSTSSQTAWALLGLMAAGEVDHPAVARGIDHLLGTQREHGLWDETRFTATG
FPRVFLRYHGYRKFPLWALARYRNLKRANATRVTVGM

>seq\_ID 27
MAHQETMASETSISLHTLACDATKLAGTYALRQVREDGHWYGEMKSNATITAEYVFLAQLGF
SIEEDRDDLIKYFLSEQNTDGSWSLAYDFPGDVSVTAEAYFALCLLGLDRSHPMASAREFTLS
KGGIAKVRVTRMFACFGLFPWSAVPELPAELILLPAAAPMSIYQLASWARATVVPMLVIRHH
RFIYALPNGRSSSNEYLDELWVDPTDKMVPYSPSLWSLWDDLTAFGPTLADNLLKALGGLRW
FPRSRIALRHCVAWILERQEPEGDIIGFIPPLHAALFALALEGYGLESPVRRGIDALQNTYAWR
DSTGLRIQGCISPILDITLMTIGLIDSSLPAESPLVARSSRYLKAHQQLGNEGDRVYNGNVPSG
GFNFYFNSWYVDIDDTAAAILAMVKQDPNLLDLGPILSAVQWILGLQNDGGWAAFDRENNY
LFLNKIPFSDMDSFCDPSTADVTRGVI ECFGLNGKNPIPRFFIDDMSATERAIDFLSTEQEADG
SWYGRWGSNYIYGTSAVLCGLVYHLEGWDDTYVMEKRKHVDTHAALDWLKRHQNPDDGGW
GERLESYYEPRLAGNGPSTASQTAWALMGLLAYLAPTDESITRGIQYLSRTQIKGELAGSWEKE
DHYTGTGFPNHFYLYCYTLYSQYFPMMLALGRYTSLSGYRPLENLESTVEDHKGNSSDC

>seq\_ID 28
MMTLREEGHKEGITPGKEQLTSDIEHSLKLATEYALSSIRSDGHWCGELRSNVTITAEYIFLRHA
LGLDLRTDAAAYCRYILSQQNCDSWGLAPYEPGDVSTTEAYLAKLLGTSPDMPAMQOAR
AFVRKAGAGAEKRVVTRIFLATFGLFPWDVAPQLPVELILLPSSCPINMYTLASWARGTIAPLLII
CHHQPVYALPEDYLDLWLDPTDKNVPYGSRLRDLRSRGLITGLAFSVVDNLLYLNGLSVPL
LRSYARRKCIQWILERQEPTEGDWAGI FPPMHAS IYAFVLEGYELNDPVRVLRG IQALENFAWEDE
KGRKIQAQCVSPVWD TALMS IGLCDAMS PDKQILQQAITWIRNRQLLKP CGDWIRIYR SKLAPGFE
SFEYENSHYPDIDD TAAIILAQLKQDPQSVASDSVIAAATWILGMQNPDDGGWAAFPDVENDKLFL
NKIPFSDMDSLCDTSCADI TGRILEAFGLMMKRELKRPVLSPLMRHACIRGITTYLASTQESNGA
WFGRWGCNYIYGTCHALGLVA PALQWLKSKQNDGGWGEPLLSYRTPGTQLQQQSTPSQTA
WALMGLLAHLPLTDPAIERGI RVLVCSQQPEKNGASWPEAVYGTGFPNHFYLYGYDYRHY
FPMMLALGRYLQASQAQA

>seq\_ID 94
MNDLTDMATLSAGTVPAELDAVARATDALLAAQNADGHVVELEADSTIPAEYVLLVHYLGE
TPNLELEQKIGRYLRRIQADGGWPLFTDGPANI SASVKAYFALKVIGDDENAEHMQRRARRAIH
AMGGAEMSNVFTRIQLALYGAIPWRAPMMPVEIMLLPQWPPFHL SKVSYWARTVIVPLLVLNA
KRPLAKNPRGVRIDELFDPPVNAAGLLPRQGHQSAGWFAFRAVDHVLRAVDGLFPAYTRERAI
RQAVAFVDERLNGEDGLGAIY PAMANAVMMYDVLGYAEDHPNRAI ARKSI EKLVLVHVEDEAYC
QPCLSPVWDTSLAAHALLETRDPRAEQAAVRGLDWLRPLQIILDVVRGDWISRRPHVRPGGWAF
QYANPHYPDVDDTAVVAMAMDRVAKLQNSDQTYRESIARAREWVVMQSSDGGWGFEPEN
TQYYLNNIPFSDHGALLDPTADVSGRCLSMQLGETALNSDAARRALDYMLKEQEPDGSW
YGRWGMNYVYGTWTALCALNAAGLGPDDARVKRAQWLLSIQNKDGGWGEDGDSYKLNRYR
GYEPAPSTASQTAWALLGLMAAGEVNNPAVKRGI DYLI AEQKEHGLWDEARFTATGFPFVFLY
RYHGYRKFPLWALARYRNLKRDNTRTVGM

>seq\_ID 30
MERSLLVPASIDSHSRESETGLDQAIVRARAALLGRQGDGHWCFELESDCITITAEYILMMH
FTDEIDEDLQERMARYLRATQVQETHGGWPQYVGGAILDLSCTVKAYYALKAGDSPEAPHMR
RAREAVLALGGAASNVFTRIQLALYGVVWPYAVPMPVEIMLLPQWPPFHL SKVSYWARTVIVPLLVN
PLTILCSLKRANPRKRVIREL FVTAPEQERHYFLRGGLNRIPLGLDKFARTLDRWMPKSLRQ
HAIRKAEAWFLPRMNGEDGLGAI FPPMVNCEAMILLGYPKDHDPARKTCLRSIQKLI VHRDDGS
AYCQPCVSPVWD TAWSAMALIHSGDDTATQTAIARAGDWLVQRQELDCRGDWEAQAQPAAP
GGWAFQYANGYYPIDDTALVAALLHISDRRRGQPGQHAFNIDRAVDWMLALQSRNGGFAAF
DADNTHYYLNAIPFADHGALLDPTEDVSGRVAACLGI LKRDQDRDGLRRCIDYLRRTTQQPDG
SWWGRWGSNYIYGTWSALSGLALAGEDLRQPYLRKSVDLWRTRQHPDGGWGETNDSYIDP
HLAGTNAGISTPHSTAWAVLAQLAMGEVESDSVRRGIAFLACQQTDLGLWSHPSHNAPGFPR
VYLYKHGYAAFPYLYALARYRHLLNRSREQR

>seq\_ID 98
MNDMTEMHTLDATAAPAGLDAVARATDALLAAQQADGHVVELEADSTIPAEYVLLVHYLGE
APNVELEQKIARYLRRIQPDGGWPLFTDGPANI SASVKAYFALKVIGDDENAEHMQRRARRAIH
AMGGAEMSNVFTRIQLALYGVVWPYAVPMPVEIMLLPQWPPFHL SKVSYWARTVIVPLLVN
AKRPAKINPRGVRIDELFKGAPVSTGLLPKQPHQSAGWFAFRAVDGVLRLVDGLFPYTRER
AIRQAVAFVDERLNGEDGLGAIY PAMANAVMMYALGY PEDHPNRAIARRSIEKLLVVGQEAY
CQPCLSPVWDTSLAAHALLETRDPRAEQAAVRGLDWLRPLQIILDVVRGDWISRRPHVRPGGWA
FQYANAHYPDVDDTAVVAMAMDRVAKLDRDAYRESIARAREWVVMQSSDGGWGFEPEN
TQYYLNNIPFSDHGALLDPTADVSGRCLSMQAQFGETSASSEPARRALDYMLKEQEPDGSW
YGRWGMNYIYGTWTALCSLNAAGLGHDDPRVKRAQWLLSIQNDGGWGEDGDSYKLDYR
GYERAPSTSSQTAWALLGLMAAGAVDNPAVARGVDYLLGTQREHSLWDETRFTATGFPFVFLY
LRYHGYRKFPLWALARYRNLKRANATRVTVGM

>seq\_ID 187
MTSDTASAAALDPRRLATSITRASRALHDVQQPD SHWVFELEADVITIPAEYVMMRHYFAEPVD
AEIEAKIAKYLRMNQNDGGWSLFYGHFEDMSASVKAYYALKMIGDSDPAPHMKKAREAMLA
RGGASRANVFTRIIMLALFGQVSWKAVPMPVEIMLLPRWFPFHLTKVSYWARTVIVPLLVMLT
LKPRAKNPRGIVRELFLEDPTQTVGPTPKAAHQSLWFTSFDI IDRVLRITDPPFPKGMKRKRAI
KAEAFVTERLNGVDGLGAI FPA MVNSIMMYDVLGYPPNDPNRALARESVERLLVIKDD EAYCQP

## Enzyme Sequences

CVSPVWDTALAHSMLSEGEADIEAAKAGLDWLLPRQVLDLKGDWADKRPDVRPGGWAFQ  
 YNNAHYPDLDLDTAVVVMAMDRVRRLDGTTKYDEAIARATEWILGLQSENGGWAAPFDADNLEY  
 YLNNIPFADHGALLDPPTEDEV TARCLSMQLAQLGDTLETSEPMRRGVEYLRKTLQPDGGSWFR  
 WGINVYGTWVLCALNAVGVPHDDPMIAKAADWLESIQNEDGGWGEDGNSYKLNKYGER  
 AATTASQTAWATLALMAAGRVRDATQRGIDNLVQSQEADGFVWGPYTTGGGPRVFLRY  
 HGYSKFFPLWAMARYRNLRSNSRFVAGM

>seq\_ID 207

MNKHSGNRTAIDPAALEMSIASATEALLAYRHADGHWAFFLEADSTIPSEYILLRHYLAEPIDVVL  
 EAKIGNYLRRTOGAHGGWPLVHDGPFDMASVKSYPALKMIGDSVDAAHMVKAREAIRARGG  
 AANSNVLTRFLALALYGVVSWRAVPVLPVLEIIVLLP I WSPFHLYKISYWARTTIVPLMVLAFLKPRK  
 NPKGVGIEELFLQDTKSVGMNPKAPHQSWGWFLFRGIDGILRVI EPHLPKCLRERAIASALAF  
 EERLNGEDGMGAIYPSMANIVMMYDALGKDDHFPRAIARRAIDKLLVI GEEEAQCQCLSPVW  
 DTAL TCHALQEVGGANAVAKAKQGLDWLKPRQVLDVKGDWAVKAPNIRPGGWPFQYNNAHY  
 PDLDDTAVVVMAMDRVRRHAGSKEYATAIARGREWI EGMQSRDGGWAADFVNNLEYYLNNL  
 PFADHGALLDPPTEDEV TARCVSMLAQVGEFTQRSKAVAEGIAYLRRTQHAEGSWYGRWGLNY  
 IYGTWVLCALNAAGIDHQDPMIRKAVEWLVSIQSWDGGWGEDAISYRLDYSGVEQAPSTSSQ  
 TAWALLGLMAAGEVEHPAVARGVNYLKNQAQTENGLWDEQRYSYATGFPFRVFLRYHGYSKFFP  
 LWALARYRNLRSNTN

>seq\_ID 29

MTTGHRRQDDGLSERERLIHEAGLTLQRSMDYAYNVVRS DGHWC GEMSSNVTI TAEYIFLRQA  
 LGLDLKTDGAAYCRHILSQNSDGSWGLAPEYPGDVSTTEAYLALKMGLSTDPAMQQAK  
 APVNLNAGGVAKVRVFRIFLATFGLFPWKAVPQLPVELILLPSACPINIYKFSWARGTIAPLLIIC  
 HHQPVYALPNGVFAENEYLDLWQDSTNKSEFYSIWELLSQDITGLTFLSLDKLLYQLNGL  
 RSIPLLRSYALKQCKMKWILERQEP TGDWAGIFPMPHASVYAFVLEBGYKLEDPVRLGIBALENF  
 AWEDAKGRVQPCVSPVWDTTLMSTALS DAATPNHQIVDRATQWIRDRQLLEPRGDWRVYRP  
 RLAPGGFSEFYTNHYPI DDDSAAILAQVKHDPISANSSSVIAAATVILGMQNPDGGWAADFV  
 ENDKLFLNKIPFSDMDSLCDTSCADITGRILEAFGLLIRRVDPDKSSQLFQLLPAIRAACRRGIRY  
 LASTQEANGAWFGRWCNYYGTSHALCGLAYFLQEDQVPMVQPALQWLKSSQNDG  
 WGESLLSYQS PERKQRSTASQTAWALMGLLAHLPHDTIVIERGIRWLVSQRVETLGTSTW  
 EPVYTGTFPNNHFYLYGYDYRHYFPMMLGRYLRGVQG

>seq\_ID 25

MLQTEAITTEGLRVRSLSPDDPLLPRIKQAIKLSGQHSRGMHSDGHWCGEVKTNATTSAEHV  
 LLCQALGINLDADREAFISWFRCTQGADGGWSTAPDQAGDISVTV EAYLALKILGLESDDAAMR  
 RARDFAIAGGVAKVRIPTRIYALFGLFPWAAPPELPELILLPSRVPVSIYHWSAWARATVVP  
 LLIISHHRPIYALPGGGKMTSSDYLDLWQDSTNKSEFYSIWELLSQDITGLTFLSLDKLLYQLNGL  
 DGLRSPNPFRRFALQKCVDWILEHQEDMGDIGDIMPLLHGAMLALRLEGYP LHS GPIHRGLEAI  
 ERFAYRDKQKRIQTTFYSAFWDTSMLIALGADGAMSKPWLTRSLGWLQHQRLGNYGDK  
 VNNHGLKAGGFSFGYFN TWYDPVDDTASAVLAMIRQDERLVHSAVLDALNWLGMQNTDG  
 GWGAPDRDNDKHLFNKIPFSDMDALCDPSTPDVTGHVLEAFGLFLALS KADALADRVAASRR  
 AIRYLSDTHVLSRGWYGRWCNYYGTSAVLCALAYFGSENDALSGVRVMKDAINQAI RWLET  
 VQNPDGGWGETVDSYKDP SRAGSGPSTASQTAWAIMALLPYLPPSTEVIQRGMEYLLRQT  
 TASQGATWHEKAYTATGPPKYFMYGSLYAHYFPMMLGRYAYPCPAWHENWRLKRD

>seq\_ID 97

MNDLSQAQPLDAI LPDFADAAPSAPAPAVTGEAPTASLDAATIRATEAILAAQKPDGHVVELE  
 ADATIPAEYVLLVHYLGETP NLELEBQKIARYLRR IQLPDGGWPLFTD GALDISASVKAYFALKMIG  
 DPADAEMHVRAREAILAHGGAETVNVVFTRI L LALFGLVSWRAVPMPVPEIMLLPMWPFPHLSK  
 VSYWARTVIVPLLV LNAKRPVARNPRRVIDELFRGAPVNTGPRDRAPHQHAGWFRFFSGVD  
 VLLRAVDGLFPKSTRERAVRQAVAFVDERLNGEDGLGAI F PAMANSVMMYDVLGYPADHPNR  
 AIARQS IDKLV I KDD EAYCQCLSPVWDTSLAAHALLETGEAHEAQAERGLAWLRPLQLLDVR  
 GDWI SRPNVRPGGWAFQYNNAHYPDVDDTAVVVMAMQRSATVTSQSDVDRDAIARAREVV  
 GMQSDGGWGAPEPENTQYLYNNIPFSDHGLDPP TADVSGRCLSMQLAQLGELPQNS EPAQ  
 RAFDYMLKEQESDGSWYGRWGLNYIYGTWVLCALNAAGIDHQDPMIRKAVEWLVSIQSWDGGWGEDA  
 VSYRLDYSGVEQAPSTSSQ TAWALLGLMAAGEVENHEAVARGVAYLEREQREHG  
 LWDETRFTATGFPFRVFLRYHGYSKFFPLWALARFRHLKRNLTRVAVGM

>seq\_ID 176

MNSVNAVAPIDDAALGSGI GAATRGLLDLKQPDGHFVFELEADATIPSEYVLLRHYLGEVDA  
 ALEAKIAYVLRRIQGAHGGWPLVHDGPFDMASVKSYPALKMIGDSIDAPHMARAREAILSRGG  
 AANVNVFRFLLSLFEVL TANRSAPVLPVLEIMLLPMWSPFHINKISYWARTTMVPLMVLAALKPRA  
 RNPRGIGIRELFLQDPAVGTGPKRAPHQSPAWFTLFNSLDWILRKEIPLFPKRLRARAIEKAI AFV  
 EERLNGEDGLGAI F PPMVNTVMMYDALGFPPEHP PRAVARRGIDKLLVI GKDEAYCQCPSP I  
 WDTALTCHALLEAGSPEALSGAGKSLDWLLPKQELVLDKGDWAVKRPDVRPGGWAFQYANAH  
 YPDLDDTAVVVMAMDRVRRNDRSDKYNEA IARGREWI EGMQSRDGGWAADFADNLEYYLNNI  
 PPSDHAALLDPPTEDEV TARCVSMLAQLGETVRRSSPSMAAGVDYLRRTQLKEGWSYGRWGLN  
 IYGTWVLCALNAAGIDHQDPMIRKAVEWLVSIQSWDGGWGEDA VSYRLDYSGVEQAPSTSSQ  
 TAWALLGLMAAGEVENHEAVARGVAYLEREQREHG LWDETRFTATGFPFRVFLRYHGYSKFFP  
 LWALARYRNLRSNTNSKVVGVGM

>seq\_ID 210

MDSGTFNPGGERGNTLDASIDAARAALLGYRRDDGHVFELEADCTIPAEYVLLRHYLGEVDA  
 AALEAKIAYVLRRIQGAHGGWPLVYDGEFDMASVKSYPALKMIGDSIDAPHMARAREAILSR  
 GGAVHANVFRFLLSLFEVL TANRSAPVLPVLEIMLLPMWSPFHINKISYWARTTMVPLMVLAALKP  
 RAVNRLGVGLDELFLQDPKSIMGARGPHQNRGLFALF GAIDAVLRVIEPLIPKCLRERAIASALAF

-continued

Enzyme Sequences

AFVEERLNGEDGLGAIYPPMANTVMYKVLGYPEDHPPRAITRRGIDLLLVIGEEEAYCQPCVS  
 PIWDTSLTCHALLEAGGAEAAQPVREGLDWLLPKQVLDLKGDWAVKAPNVRPGGWAFQYNN  
 AHYPDLDDTAVVVMALDRARRDQPSAAAYDNAIARGREWIEGMQSDGGWAAFDVNNTEYYL  
 NNI PFS DHGAMLDPPTEDVTARCVSMLAQLGETEQTSKAVARGVAYLRKTKLQPDGWSYGRW  
 GNNYIYGTWAVL CALNAAGVDHQDPAIRKAVAWLASIQNADGGWGEDVSYRLDYRGYETAP  
 STASQTAWALLS IMAAGEVDHPAVARGIEYLKGTQTEKGLWDEQRHTATGPPRVFYLYRHGYS  
 KFFPLWGLARYRNL RATNSKVVGVM

>seq\_ID 23  
 MTTGHRQFDDGLSERERLIHEAGLTQRSMYAYNVVRS DGHWC GEMSSNVTITAEYIFLRQA  
 LGLDLKTDGAAYCRHILSQNSDGSWGLAPEYPGDVSTTEAYLALKMLGLSTDAPAMQQAK  
 AFVLNAGGVAKVRVFRIFLATFGLFPWKAVPQLPVELILLPSACPINIYKFASWARGTIAPLLIIC  
 HHQPVALPNGVFAENEYLDLWQDPTNKSEFYSPIWELLSQDITGLTFSLDKLLYQLNGL  
 RSIPLLRSYALKCKMKWILERQEPTGDWAGIFPPMHASVYAFVLBGYKLEDPVRLGIEALENF  
 AWEDAKGRVQPCVSPVWDTTLMSTALS DAATPNHQIVDRAIQWIRDRQLEPRGDWRVYRP  
 RLAPGGFSFEYTNSHYPI DDSAAIILAQVKHDPISANSSSVIAAATWILGMQNPDDGGWAAFDV  
 ENDKLFLNKI PFS DMDSLCDTSCADITGRILEAFGLLIRRVDPDKSSQLFQLLPAIRACRRGIRY  
 LASTQEANGAWFGRWCNYYGTSHALCGLAYFLQEDQQVPAMVQPALQWLKSOQNDG  
 WGESLLSYQS PERKEQRSTASQTAWALMGLLAHLPHTDIVIERGIRWLVSQRPVETLGTSTW  
 EPVYTGTFPNHFFLYGYDYRHYFPMMLGRYL RGVQG

>seq\_ID 91  
 MNDLSQAHVLGAAMPETAGEAQAQAANSAAAAEASAVLAPSLDAAITRATDAILAAQKPD  
 GHWVYELEADATIPAEYVLLVHYLGETPNLELEQKIARYLRRIQLPNGGWPLFTDGAIDISASVK  
 AYFALKMIGDPVDAEHMVRAREAILAHGGAETVNVFTRILLALFGVSWRAVPMMPVEITLLPM  
 WFPHLSKVS YWARTVI VPLLVNNAKRPLARNPRVRIDELFRGAPVNTGMPARAPHQHWGWF  
 GPPRVVDTVLRVAVDGLFPKATRERAVREAVAFVDQRLNGEDGLGAIFPAMANSVMYDVLGY  
 PADHPNRAIARRSIEKLLV IKDDEAYCQPCLSPVWDTSLAAHALLETGDARAEQAERGLAWLR  
 PLQILDVRGDWISRRPNVRPGGWAFQYNNAYPDVDDTAVVAMAMHRSEALTHSGADREIA  
 RAREWVVMQSSDGGWGAFFEPENTQYYLNNIPFSDHGALLDPTADVSGRCLSMALQLGEP  
 PQNS EPAQRALDYMLKEQADGSYGRWGLNYIYGTWTALCSLNAAGLPHDDPRIRRAAQW  
 LLSIQNEDGGWEGGESYKLDYRGYERAPSTASQTAWALMGLMAAGEVDHEAVARGIEYLQR  
 EQREHGLWDETRFTATGPPRVFYLYRHGIRKFFPLWALARYRHLKRNLTRVAVGM

>seq\_ID 213  
 MDSGSYTTGVERNALEASIDAARSALLNRRDDGHWVFELEADCTIPAEYVLLRHYLGEVDA  
 ELEAKIAVYLRRIQAGAGGWPLVHDGDFDMSASVKGYFALKMIGDSIDAPHMVRAREAIRSRG  
 GAIHNSVTRFLLTLYGVTWRRAVPVLPVEIMLLPSWSPTLTKISYVVARITMVPVLLVLCALKPKQ  
 AKNPKGVGIDELFLQDPKTI GMPVKAPHQNWALFKLFGSIDAVLRVIEPVMPKGIKRAIDKALA  
 FTEERLNGEDGMGAIFPMANAVMMEALGYPEDYPPRASQRRGIDLLVDRGDEAYCQPCVS  
 PWWDTALASHAVLEADGHEGAKSVRPAIDWLLPRQVLDVKGDWAVKAPNVRPGGWAFQYNN  
 AHYPDLDDTAVVVMALDRARKDQPNPAYDAAIARAREWIEGMQSDGGWGAFFDINTEYYLN  
 NIPFSDHGAMLDPPTEDVTARCVSMLAQLGETMDSSPALARAVGYLRDTQLAEGSWYGRWG  
 MNYIYGTWAVL CALNAAGVPHADPMIRKAVAWLESVQNRDGGWGEDAVSYRLDYRGYESAP  
 STASQTAWALLALMAAGEVDHPAVARGIEYLKSTQTEKGLWDEQRHTATGPPRVFYLYRHGYS  
 SKFFPLWALARYRNLQATNSKVVGVM

>seq\_ID 196  
 MSMTSREDHDASSLSIQVEHALKLSNDYALGLVHPDGHWYEMNTNVTVAEYVFLRQALRL  
 DLKTDIAAYCHYLLSQNSDGSWGLAPEYPGDVSTTEAYLALKMLGTSPTAMPNRARAFVLK  
 AGGIARVRIFRIFLATFGLFPWSAVPELPELMLLPSICPINIYKFASWARGTIAPLLIICHHPQVY  
 SLPNKSTDNDDYLDLWVDTCTNKSPYGLPLWDLMSQGEFAGLAFGVLDKVLQNLGRLSIP  
 ITRAYARKQCIQWILERQEKTGDWAGIFPPMHANMYAFTLEGYKLDLDDVPVRLGQALERFAWED  
 EKGKRIQACVSPVWDTALMTI GLCDAMSPNKQTI DHALAWIRARQLEPRGDWRVYRPQLAPG  
 GFSFEYENSWYPDVDDTAAIILAQVKHDNGSISGNSVIAAATWILGMQNPDDGGWAAFDVENDK  
 LFLNKI PFS DMDSLCDTSCADITGRILEAYGLMMKPYFAKSDADPLLHTLRAACMRGMHYLAS  
 TQEPNGSWYGRWCNYYGTSHVLCGLAYFVEKRLVCMVKSALQWLKSRQNDGGWGES  
 LLSYQSPDREQOASTPSQTAWALMGLLSHLPTDDAIIERGIRYLVSQRPEKIGSSWPQAEY  
 TGTGFPNHFFLYGYDYRHYFPMMLGRYLQGSRGLN

>seq\_ID 99  
 MNDLSQTPLAAVLPEAADAPAVADASATAAPEPVQAASPSALDASITRATDITLAAQKPDGH  
 WVYELEADATIPAEYVLLVHYLGETPNLELEQKIARYLRRIQLPNGGWPLFTDGAIDISASVKAY  
 FALKMIGDPVDAEHMVRARDAI LAHGGAERANVFTRILLALFGVSWRAVPMMPVEIMLLPVWF  
 PFHLSKVS YWARTVI VPLLVNNAKRPLARNPRKVRIDELFRAPVNTGMNERAPHQHWGWF  
 FRCVDTVLRVAVDGLFPKATRERAIRAAVAFVDERLNGEDGLGAIFPAMANSVMYDVLGYPAD  
 HPHRAIARKSLDKLLV IKDDEAYCQPCLSPVWDTSLAAHALLETGEARAEQAERGLAWLRPL  
 QILDVRGDWISRRPNVRPGGWAFQYNNAYPDVDDTAVVAMAMHRSAALTQSDVDREAIARA  
 REWVVMQSSDGGWGAFFEPENTQYYLNNIPFSDHGALLDPTADVSGRCLSMFAQIGELPQS  
 SEPARRAFDYMLQEQEPDGSYGRWGLNYIYGTWTALS SLNAAGLPHDDPRMRRAAQWLV  
 SIQNEGGWEGGESYKLDYHGYERAPSTASQTAWALLGLMAAGEVNHAEVARGIDYLRQRE  
 QREHGLWDETRFSTATGPPRVFYLYRHGIRKFFPLWALARFRHLKRHLTRVAVGM

>seq\_ID 85  
 MIRRMNKSAPSPWALSADAAIARGRDALVRLQPDGWSWCFELES DATITAEYILMMHFMDRIDD  
 VRQERMARYLRANQRDLTHGAWDLVYDGPADVSCSVKAYFALKKAAGDSEHAPHMIRARDAIL  
 KLGGARSNVFTRILLALFGQVPWRAAPFMPPIEFVLFKQWVPI SMYKVAWYARTMVPVLLVCS

-continued

Enzyme Sequences

LKARARNPRNVSIRELFVTPPEQERHYFLPARGMRRFLALDRTRVRIEPLLPKRLRQRAIRHAE  
 AWCERMMGEDGLGGIFPPIVYSYQMMQVLYGYPDDHPLRRDCENALEKLLVTRPDGSMYCP  
 PCLSPVWDTAWS TMALEQARGVAAPETGDTASGALRELDER IARAYDWLATRQVNDLRGDWI  
 ENAPADVPEGGWAFQYANPYYPDI DD TALVTAMLDRRGRTHRAGDGHYASRVARALDWM  
 RGLQSRNGGFAAFDADCDRMYLNAI PFADHGALLDPPTEDVSGRVLLCFGVTKRAADRASLAH  
 AIDYVKRTQQPDGSSWGRWGTNYLYGTWSVLAGLALAGEDKSQPYITRALDWLRARQHADG  
 GWGETNDSYIDPKLAGTNDGESTSNCTAWALLAQMAFGDCSDSVKRG IAYLQSVQQEDGF  
 WWRSHNAPGFPRIFYLKYHGYTAYFPLWALARYRRLAGAKDADATRSPASATPATDNALA

>seq\_ID 93  
 MIRAMNKSALSPWSALDTAIARGRDALRQLQPDGSGWCFELES DATI TA EYILMMHFMDRIDDA  
 LQERMARYLRAIQRLDTHGAWDLVYDGPDPVSCSVKAYFALKAAGDSEHAPHMIRAREAI LKL  
 GGAARSNVFTRILLATFGQVPWRATPFMP IEFVLPKWPVISMVKVAYWARTTMVPLLVLCSLK  
 ARARNPRNVAIPELFVTPPDQERHYFPTRGMRRAFLLIDRVVRHVEPLLPKRLRRRAIRHAE  
 WCAQRMNGEDGLGGIFPPIVYSYQMMQVLYGYPEDHPLRRDCENALEKLLVTRPDGSMYCP  
 LSPVWDTAWS TMALEQARSVAVPESEDESARALDEL DAR IARAYDWLATRQVNDLRGDWIENA  
 PADTQPGGWAFQYANPYYPDI DD SAVVTAMLD RRRGRTHR NADGSHPYAARVARALDWMRAL  
 QSRNGGFAAFDADCDRLYLNAI PFADHGALLDPPTEDVSGRVLLCFGVTRRAEDRASLARAID  
 YVKRTQQPDGSSWGRWGTNYLYGTWSVLAGLTLAGEDPSQPYIARALEWLRARQHADGGW  
 GETNDSYLDPALAGTNGGESTSNCTAWALLAQMAFGDCASDSVKRG IAYLQSVQQEDGF  
 HRSHNAPGFPRIFYLKYHGYTAYFPLWALARYRRLAGAAEARARASSGRAPHAADTALA

>seq\_ID 168  
 MGKVELTHRMSTQDITLDDVERRVSLASKALMRLAGPDGHWC FELEADATIPSEYILYHHFRG  
 STPSAELEGKIANYLRRRQSAQHDGWSLVHDGPPDMSATVKAYFALKMIGDSIEAPHMRRARE  
 AILRRGGAHANVFTRILLALYGEVPSAVPMPVEVMLLPWFPHLDKVS YWARTVMVPLF  
 VLQAKKPRARINPRGIGIQELFVPEPERVKRVPAGQES SPWRPVFAADKVLQKVEGSPFAGS  
 RARAIDKAVAFVFERLNGEDGLGAI FPAMVNAVLMY EALGYPEDHPLVATARSVEKLVTVKEH  
 EAYVQPC LSPVWDTALS AHALMEAGGVEAERHAKRALDWL KPLQVLDIKGDWAASKPNVRPG  
 GWAFQYANPHYPDLD DTA VVMAMDR AQVRRSPGDAADYQQS IARAREWVEGLQSRDGG  
 WAAFADANTYHYLNYIPFSDHGALLDPPTADVTARCVSMLAQLGETRESCPLDRGVAYLLD  
 QEADGWSYGRWGMNYIYGTWSVLCALNAAGVDPASEPVRRAVNWLT TI QNPDGGWGEDAA  
 SYKLEYRGYERAPS TASQTAWALLGLMAGEADS PAVARGINYLTRS QGADGLWTE DRYTAT  
 GFPRVFLRYHG YAKFFPLWALARYRNLQQSNSRRVAVGM

>seq\_ID 184  
 MKKFGGMARTSLQAQSPGNNTPSMDEKMLKAGLEAARGALLAQ QREDGHWCFLEADCTI  
 PAEYILMMHFMDVLDLLEVR IARFIREKQDV AHGGWPLYGGEPDLSCSVKAYALKIVGDS P  
 DAPHMVRARAAILKHGGAARANVFTRILLALYDQLPWRGVPFVPEIILFPKWFPHFTSKVAY  
 WSRVTVMVPLSILCSL KARAANPRKVAI RELFTVPPGEERNYFPVTRALNRVFLLIERTLSLEPFI  
 PQGVRRALARAESWIVERLNGDSGLGAI FPAMVNAGEALALGYPYDHPAREQCRKALRLL  
 VEEGERTWCQPCVSPVWDTVLTCLAFQEDTEVDQKPIRKALDWLVPCQVLDAPADWQEDHP  
 GLPGGGWAFQYANPHYPDLD DTA AVAWALYQADPKAYQES I SRAADWLAGMQSSNGGFAAF  
 DSDNTYYLNEIPFADHGALLDPPTSDVSARCAGFLALYGGSRHKQALERSLAYLFNEQEASG  
 AWFGRWGSNYIYGTWSVLEAFRLAGIDAGHPAIRRAVHWLKS VQRDEGGWGESNDSYLSQP  
 QAGQFHTS TS PHTAWALLALMGAGEWR.SHEVHRGIAYLLRBDSDGLWHEPWF TAPGFPRV  
 FYLKYGYTKYFPVWALTRFHALNRKFPG

>seq\_ID 12  
 MMYNNQWYFNQFNDIFCFPEQKKEYFPPTGTNISLNLKRPDRQLLAHGASDLNGPFHLSQH  
 NAFSAMLAEVQKVLRLAVGHS LQLR TDGAWCGEVHSNATFTAQYVFLQQQLGLPLDPTBIE  
 GLSRWLF SQQNEGDSWGLGPGGLGGDVSTTETYLALKILG VSPEDPRMAARSSIKAGSLPA  
 TRMFTRVFLASFLIPWSAVPPLPAELI LLPTLPFVNIYNLSSWARATCVPLLLIRHHEPLHSLPN  
 GRHAENFDLDELWTKDIPRDFCYTTPLSRMWRLGDYAGIFPFSADHGLRFLGQYFNSPLRNL  
 RRKIINWIDLHQEQSGEWAGYWPQHNNI WALSLDEGYSLDHPLVLRGIAAVKSFVLHDVTGMR  
 AQVTVSQVWDTALMSIALSDSAPSTGIISPTQAI DWMHHEVASHRGDWRVLRPKLATGGPCF  
 EEFNTLYPDVDDTAAVIMALI KSNPAHLISGCVRRAAQWILGMQNRDGGWGFADWNNDKFPLN  
 KIPFSDMDSLCDPSTPDVTGRIIECFGMMAGRHGYSLDGPLESRLRASSQLAIAYLLGCQENN  
 GSWGRWGVNLYGTSNVLCGLAYYDRSGLSKDGGKSNSHIVSAVDRASEWLKARQHSN  
 GGWGEPEYSDAQLAGCQPTASQSAWVTMALLNLYLSP TDEVIQRGISYLVR SQVKYGDSE  
 RATWPLERYTATGFPGHLYMEYDYRHYFPIMALGRYVKNKLSSESKLL

>seq\_ID 100  
 MIRMTTPTPSPWSALDTAIARGRDALVRLQPDGSGWCFELES DATI TA EYILMMHFMDKIDDL  
 RQEKMARYLRANQRLDTHGGWALVYDGPDPVSCSVKAYFALKAAGDSEHAPHMVRARDAI LK  
 LGGAARANVFTRILLATFGQVPWRAAPFMP IEFVLPKWPVISMVKVAYWARTTMVPLLVLCSL  
 KARARNPRNVSIRELFVTPPEERQYFPARGMRKFLALDRTRVHRHVEPLMPKGLRQRAIRHAE  
 AWCERMMNGEDGLGGIFPPIVYSYQMMQVLYGYPDDHPLRRDCENALEKLLVTRPDGSMYCP  
 CLSPVWDTAWS TMALEQARGVAVAEDGEPGDARRALDERI TRAYDWLAERQVNDLRGDWIE  
 NAPADVQPGGWAFQYANPYYPDI DD TAVVTAMLD RRRGRTHANADGTNPYATRVARALDWMR  
 GLQSRNGGFAAFDADCDRLYLNAI PFADHGALLDPPTEDVSGRVLLCFGVTKRAEDHASLAR  
 CIDYVKRTQQPDGSSWGRWGTNYIYGTWSVLAGLALAGEDKSQPYIARAEWLRARQHADGG  
 WGETNDSYIDPKLGGTNGGESTSNCTAWALLAQMAFGDCSDSVKRG IAYLQSVQQEDGF  
 WWRSHNAPGFPRIFYLKYHGYTAYFPLWALARYRRLAGVANKRVSTADKTADAMA

-continued

Enzyme Sequences

>seq\_ID 84  
MIRRMNQSAPSSWSALDAAIARGRDALVRLQQPDGSGWCFELESDATITAEYILMMHFMDRID  
VRQEKMARYLRLANQRDLTHGAWDLVVDGAPDVSCSVKAYFALKAAGDSEHAPHMIRARDAIL  
KLGGAARSNVPTRI LLATFGQVPWRAAPFMAVEFVLPKWPVISMVKVAYWARTTMVPLLVLC  
SLKARARNPRNVSIRELFTVTPPEQERHYFPPARGMRRLLFALDRTRVRIEPLLPKRLRQRAIRH  
AEAWCAERMNGEDGLGGIFPPIVYSYQMMQVLGYPDDHPLRRDCENALEKLLVTRPDGSMYC  
QPCLSPVWDTAWSTMALQARGVAAPETGDTATGAPRDLDRGRIARAYDWLARTQVNDLRGD  
WIENAPADVEPGGWAFOYANPYYPIDDDTALVTAMLDRRGRTHRAADGTHPYASCVSRALDW  
MRGLQSRNNGFAAFDADCDRMYLNAIPFADHGALLDPPTEDEVSGRVLCCFGVTKRAADRASL  
ARAIYVKRTQQPDGSGWGRWGTNYLYGTWSVLAGLALAGEDKSPYIARALDWLRARQHA  
DGGWGETNDSYLDPLKAGTNGGESTSNCTAWALLAQMAFGDCESDVKRGIAYLQSVQQED  
GFWWHRSHNAPGPPRI FYLYKHGYTAYFPLWALARYRRLAGAKDAGATRS GASASATSVTD  
DALA

>seq\_ID 86  
MIRRMNKSAPSPWSTLDTAIARGRDALVRLQQPDGSGWCFELESDATITAEYILMMHFMDRID  
VRQEKMARYLRLANQRDLTHGAWDLVVDGAPDVSCSVKAYFALKAAGDSEQAPHMIRARDAIL  
KLGGAARSNVPTRI LLATFGQVPWRAAPFMPPIEFVLPKWPVISMVKVAYWARTTMVPLLVCS  
LKARARNPRNVSIRELFTVTPPEQERRYFPPARGMRRLLFALDRVVRHIEPLMPKRLRQRAIRHA  
QAWCAERMNGEDGLGGIFPPIVYSYQMMQVLGYPDDHPLRRDCENALEKLLVTRPDGSMYVY  
PCLSPVWDTAWSTMALQARGVAAPETGETAAGTLRELDERIRARAYDWLAAQVNDLRGDWI  
ENVPADVEPGGWAFOYANPYYPIDDDTALVTAMLDRRGRTHRHADGTHPYAFRVARALDWM  
RGLQSRNNGGFAAFDADCDRMYLNAIPFADHGALLDPPTEDEVSGRVLCCFGVTKRAEDRASL  
ARAIYVKRTQQPDGSGWGRWGTNYLYGTWSVLAGLALAGEDKSPYIARALDWLRARQHADG  
GWGETNDSYLDPLKAGTNGGESTSNCTAWALLAQMAFGDCESDVKRGIAYLQSVQQEDGF  
WHRSHNAPGPPRI FYLYKHGYTAYFPLWALARYRRLAGAAAAPPALVAADTALA

>seq\_ID 80  
MIRRMNKPAPSPWSALDAAIARGRDALMRLQQPDGSGWCFELESDATITAEYILMMHFMDKIDD  
ARQEKMARYLRAIQRLDTHGGWDLVLDGDDPDLSCSVKAYFALKAAGDSEHAPHMVRARDAIL  
KLGGAARSNVPTRI LLATFGQVPWRATPFMPPIEFVLPKWPVISMVKVAYWARTTMVPLLVCS  
LKARARNPRNVAIPELFTVTPDQERQYFPPARGMRRALDRVVRHVEPLLPKRLRQRAIRHA  
QAWCAERMNGEDGLGGIFPPIVYSYQMMQVLGYPDDHPLRRDCENALEKLLVTRPDGSMYC  
QPCLSPVWDTAWSTMALQARGVAAPETGDTATGAPRDLDRGRIARAYDWLAAQVNDLRGDWI  
IENAPADTQPGGWAFOYANPYYPIDDDSAVITAMLDRRGRTHRNADGSHPYAARVARALDWM  
RGLQSRNNGGFAAFDADCDRMYLNAIPFADHGALLDPPTEDEVSGRVLCCFGVTKRAEDRASL  
ARAIYVKRTQQPDGSGWGRWGTNYLYGTWSVLAGLALAGEDPSQPYIARALDWLRARQHAD  
GGWGETNDSYIDPALAGTNAGESTSNCTAWALLAQMAFGDCESVVKRGIAYLQSVQQDDGF  
WHRSHNAPGPPRI FYLYKHGYTAYFPLWALARYRRLAGGASSAGAHTVPASTGADAALA

>seq\_ID 82  
MNKPAPSPWSALDAAIARGRDALMRLQQPDGSGWCFELESDATITAEYILMMHFMDKIDDVRQE  
KMARYLRAIQRLDTHGGWDLVVDGDDPDLSCSVKAYFALKAAGDSEHAPHMVRARDAILALGG  
AARSNVPTRI LLATFGQVPWRATPFMPPIEFVLPKWPVISMVKVAYWARTTMVPLLVCSLKA  
RARNPRNVAIPELFTVTPDQERHYFPPARGMRRALDRVVRHVEPLLPKRLRQRAIRHAQAWC  
AERMNGEDGLGGIFPPIVYSYQMMQVLGYPDDHPLRRDCENALEKLLVTRPDGSMYVYVY  
PCLSPVWDTAWSTMALQARGVAAPETGDTATGAPRDLDRGRIARAYDWLAAQVNDLRGDWIENAP  
ADTQPGGWAFOYANPYYPIDDDTAVVTAMLDRRGRTHRNADGSHPYAARVARALDWMRGLQ  
SRNNGGFAAFDADCDRMYLNAIPFADHGALLDPPTEDEVSGRVLCCFGVTKRAEDRASLRAI  
YVKRTQQPDGSGWGRWGTNYLYGTWSVLAGLALAGEDPSQPYIARALDWLRARQHADGGWGE  
TNDSYIDPDLKAGTNGGESTSNCTAWALLAQMAFGDCESVRRGIAYLQSVQQDDGFWHR  
SHNAPGPPRI FYLYKHGYTAYFPLWALARYRRLASGVSSAGVHAVPASTGADAALA

>seq\_ID 108  
MNDLSQTQPRDAVLPEAAGAVPPASAPAPAAASEAPAASLDTAITRATDAI LAAQKPDGHVY  
ELEADATI PAEYVLLVHYLGETPNVLEQKIARYLRRIQLPDGGWPLFTDGAPDVSA SVKAYFAL  
KMGDPDAEHMVRAREAI LANGGAEAVNPTRI LLALFGVVSRAVPMMPVEIMLLPMWPPF  
HLSKVS YWARTVIVPLLVNNAKRLARNPRRVRIDELFRGAPVNTGPRDRAPHQHAGWFRFFS  
GVDMLLRAVDGLFPKATREAVRAAVAFVDERLNGEDGLGAI FPMAMNSVMMYDVLGYPADH  
PNRAIARQSI EKLKLVKDD EAYCQPCLSPVWDTSLVAHALLETGEAREAEQAERGLAWLRPLQIL  
DVRGDWISRRPNVPRPGGWAFOYANPYYPIDDDTAVVVMAMHRSAAALTHESEVDREAIARAE  
WVVMQSSDGGGWAFAFEPENTQYLLMNI PFS DHGALLDPPADVSGRCLSMALQGLGELPQGS  
EPAQRAFAYMLKEQEPDGSWYGRWGLNYI YGTWTALCSLNAAGMPHDDPRMKRAAKWLLSI  
QNEGGWEGGESYKLDYHGYERAPSTASQTAWALMLMAGEVNHAEAVARGVAYLQREBQ  
REHGLWDETRPTATGPPRVFYLRYHGYRKFPLWALARFHLKRHGLTRVAVGM

>seq\_ID 169  
MREAAVSKVETLQRPKTRDVS LDDVERGVQNAARALTEMTQTDGHCIFELEADATI PSEYILFH  
QFRGTVPDRGLEAKIGNYLRRTQSKVHGGWALVHDGPPDMSATVKAYFALKMIGDDIEAPHM  
RAARKAILQRGGAANANVFTRILLALYGEVPAWAVVMPVEMHLPKWFPPHLDKVS YWART  
MVPLFVIQAKKPRAKNPRGIGVAELFVTPPDSVTRWPGSPHATWPWPITFGAIDRVLQKTDQH  
FPKVPRQRAIDKAVAWV SERLNGEDGLGAI FPMVNSVLMYEVLYGPPDPHPQVKIALEAI EKL  
V AEKDD EAYVQPCLS PVWDTAL TSHAMLETGGAAAEANARAGLDWLKPLQILDI KGDWAETKPN  
VRPGGWAFOYANPYYPIDDDTAVVVMAMDRARQHQGLVSGMPDYSAS IARAREWVEGLQS  
ADGGWAAPDADNMHHYLNHI PFS DHGALLDPPADVTARVVSMLSQGLGETRETRETRALDRGVT  
YLLNDQEKDGSWYGRWGMNFI YGTWSVLCALNAAAGVDPQSP EIRKAVAWLIRIQNPDDGGWG

-continued

## Enzyme Sequences

EDASSYKLNPEFEPGYS TASQTAWALLALMAVGEVDDPAVARGVNYLMRTQGQDGLWNEER  
YTATGFPRVFLRYHYGPKFFPLWAMARFRNLKKGNSRQVQFGM

&gt;seq\_ID 163

MREAAVSKVETLQRPKTRDVS LDDVERGVQSAARALDTMTQADGHICFELEADATIPSEYILFH  
HFRGTEPRAGLEAKIGNYLRR TQSKVHGGWALVHDGPFDMASV KAYFALKMIGDDIEAPHM  
RAVRKAILQRGGAANANVFTRILLALYGEVPTAVPVMPVEMHLPKWFPPHLDKVS YWARCT  
MVLPLV IQAKKPRAKNPRGVVAELFVTPPDSV RTWPGSPHATWPWP IFGAIDRVLQKTQDH  
FPKVPRQRAIDKAVAWV SERLNGEDGLGAI FFSMVNSV LMYEVLGYPPDHPQVKIALEAI EKL  
AEKDDEAYVQPCLS PVWDTAL TSHAMLEVGGTQAEANARAGLDWLKPLQILD IKGDWAEKFP  
NVRPGGWA FQYANPHYPLD DDTAVVVMAMDRAQRQHGLVSGMPDYSTSIARAREWVEGLQ  
SADGGWAAFDADNHHYLNHI PFS DHGALLDPPTADVTARVVSMLAQLGETRETSRALDRGV  
TYLLNDQEKDGSWYGRWGMNFI YGTWSVLCALNAAGVDPQSP EIRKAVAWLIRIQNPDGGWG  
EDASSYKLNPEFEPGYS TASQTAWALLALMAVGEVDDPAVARGVNYLMRTQGADGLWNEER  
YTATGFPRVFLRYHYGPKFFPLWAMARFRNLKKGNSRQVQFGM

&gt;seq\_ID 105

MKPNHTFSPAALDAAILRGRD TSLGQQPDGSGWCFELES DATITAEYILMMHFMKID EVRQAO  
MARYLRATQRVETHGAWDLYVDGAPDI SC SVKAYFALKAAGDSEHAPHMIRAREAI LKLGGAAR  
SNVFRILLATFGQVPWRAAPFMPVEFVLPKWPV ISMYKVAYWARTMVP LLLVLC SLRARAR  
NPRNVSIAELFVTPPDEERHYFP PAKGMRKLF LALDR TVRHLEPLLRRLRQR AIRHAEAWCAE  
RMNGEDLGGIFPP IVYSYQMMEV LGY PEDHPLRRDCE DALEKLLVTRADGSVYQPCLS PV  
WDTAWS TMALEQARGATPAAPDTQV SERELDARI ARAYDWL ATROQVNDLEGDWRENARPGT  
LPGGWA FQYANPHYPD IDDSAVVTAMLDRRGRAQARASGENPYAERV TRALDWMRGLQSRN  
GGFGAFDADCDRLYLNAIP FADHGALLDPPTEDVSGRVLLCFGVTKR PADRAAAAARAI EYVKRT  
QQPDGSGWGRWGTNYLYGTWSV LAGLALSGEDK SQPYIARALDWLRAHQHADGGWGETN  
DSYADPRLRATNYGESTSNCTAWALLAQMAFGDWQSDS VRRGIAYLLSVQQDDGFWWHRSH  
NAPGPRIFYLKYHYGTAYFPLWALAR YRRLAGQAAPSSPGPGTAATIADPAVA

&gt;seq\_ID 211

MTSGTTILGAERGR TLDASIDAARAALLGYRRDDGHVWFLEADCTI PAEYVLLRHYLGE PVDA  
ALEAKIAYYLRR TQGAHGGWPLVHDGEPDVSATVKAYFALKMIGDS IDAPHMAKAREAILARGG  
AIHVNVPTRFLLSMFGI L TWR SVPLVPEIMLLP MWAPFHLNKIS YWARTTIVPLMVLAA LKPR  
V NKLDI GLDEFLQDPQSI GMPAKAPHQSWGLFTLFGSIDAVLRVIEPLIPKLLRSY AIGRAVAFIE  
ERLNGEDGLGAI YPPMANTVM MYKVLGYGEDHPRAITRRGIDLLL VVGEEEAYCQPCVSP I W  
D TSLTCHALLEAGGAEALPVRKGLDWLIPKQVLDL KGDWAVKAPNVRPGGWA FQYNNAHYP  
DLDDTAVVVMALDRARRDQPSAAYDNA IARGREWIEGMQSDGGWAAFDVNNTEYLLNNI PF  
SDHGALLDPPTEDVTARCVSMLAQLGETAETS SALARGVAYLRKTQLAEGSWYGRWGLNYIY  
GTWSVLCALNAAGVAHQDPAMRKAVAWLASIQNADGGWGEDAVSYRLDYRGYESAPSTASQ  
TAWALLALMAAGEVDHPAVARGVEYLKGTQTEKGVWDEQR YATATGFPRVFLRYHYGSKFFP  
LWALARYRNL RATNSKVVGVGM

&gt;seq\_ID 76

MDSVNTAREAKESKISESEI LESSIASATQGV LFGQQSDGHVWFLEADCTI PAEYVLLRHYLA  
EPVDTVLEAKIGNYLRRVQGAHGGWPLVHDGEPDMSASV KAYFALKMIGDS IDAPHMVRARE  
I HARGGAIHSNVFTRFPLAMPGITANRAVPVLP I EIMLLPFWSPPHINKI SYWARTTIVPLMVI  
LAKPRAK NPKVGI DGLFQDPRS IGMTAKAPHQSMWFLFRSLDAI LRVI EPLFPKSLRKR  
TALAFS EERLNGEDGMGAI YPPMANLVMMYDALGKDENYPPRAVTRRGIDKLLVIGDDEAYCQ  
PCVSPVWDTTILTAHALLEAGGDKAGPAAKHGLDWLIPKQLEVEKGDWAVKRPDVRPGGWA F  
QYNNAYYDLD DDTAVVVMMDRMRREHGVGTGYDS AIDRGREWIEGMQSDGGWAAFDVNN  
LEYLLNNI PFS DHGALLDPPTEDVTARCVSMLAQLGETAKTSKHVADGVAYLRKTQHP EGSWY  
GRWGMNFI YGTWSVLCALNMAAGVRRHDDPMI RKAADWLASIQNKDGGWGEDTVSYRLDYK  
WEAAPS TASQTAWALLALMAAGEVDHPAVARGVEYL IATQNEKGLWDEQR YATATGFPRVFL  
RYHYGSKFFPLWGLARYRNLRTNSRVVGVGM

&gt;seq\_ID 179

MEQQPELISGGVGVAYP WDLGSAIEEAI LAARAALLAHLHPDGYWCFLEADCTI PAEYIMM  
MHYTGELEAALELKLARIYR ECQLQEGGWPLYGGAMD ISCSVKAYFALKLAGDDPEAAHMRR  
ARKAVLERGGAVNANVFTHIALALFGEI PWRGVPFMPPEI LLLPRWFPFHL SKVYSWRTVMVP  
LFI LAHKPRARNPRAIHI SELFVTD PQLTGYFKARSRLNRLFTL DALGRRIEPFI PRAVAKAL  
RRAAEWFTIRLNGEHGLGAI PFMVNSYEALELLGYAADHPLRQV R KGLRDLVVEQADRAYC  
QPCLSP IWD TALACLALQ EADRGS SAQVRHLDWLQARQLLDPGDWSEQHPSLPGGGW  
FQFRNDHYPLD DDTAIVAWAMQRASDPERYGAI RRATVWLLGMQSANGGFAAFDSNTRY  
Y LNEI PFADHGALLDPPTSDVTARVVALLSGDGEVHDRSALNRAVAF LHRBEQAEBCWYGRW  
GTNYI YGTWSVLTAL EQLGYDFNAPVWRKAVIWLKSVQRDDGGWGESNDTYLDHRPQDRQA  
DESTPFQTAWAVLALIAAGECRSPEVWRGVEYLLRHQRPDGLWYCPWTFAPGPRVFLKYH  
GYDAYFPLMALARYRNCVLDNDA

&gt;seq\_ID 81

MIRRMNKPAPSPWALDAAIARGRDALMRLQQPDGSGWCFELES DATITAEYILMMHFMKID  
ARQEKMARYLRAIQRLD THGGWDLVYDGDVSCSVKAYFALKAGDSEHAPHMVRARDAIL  
ALGGAARSNVFTRILLATFGQVPWRAAPFMP IEFVLPKWPV ISMYKVAYWARTMVP LLLVLC  
LKAHARNPRNIAIPEL FVTPDQERHYFP PARGMRRALALDRVVRHAEPLLPKRLRQR AIRHA  
QAWCAERNGEDLGGIFPPI VYSYQMMDV LGYPADHPLRRDCE NALEKLLVTRPDGSMYC  
QPCLSPVWDTAWSTMALEQARGVAVHEAGAPASALDEL DARIARAYDWLAERQVNDLRGDWI  
ENAPADTQPGGWA FQYANPHYPD IDDSAVVTAMLDRGRTRHNRADGTHPYAARVARALDWM  
RGLQSRNGGFAAFDADCDRMYLNAIP FADHGALLDPPTEDVSGRVLLCFGVTKRADDRA SLA

-continued

## Enzyme Sequences

RAIDYVKRTQQPDGSSWGRWGTNYLYGTWSVLAGLALAGEDPSQPYIARALAWLRARQHAD  
GGWGETNDSYIDPALAGTNAGESTSNCTAWALLAQMAPGDGESESVKRGIAVLQSVQDDGF  
WHRSHNAPGFTRI FYLYKHGYTAYFPLWALARYRRLAGGASSAGAHAVPASTAADAALA

&gt;seq\_ID 22

MATLTTMATTATMATTEASQPLEAQARTALTKATSYAWETIISNRHWCGELESNTVTC EHIFFL  
YVLYQHIDPDEGSQYRQWLLSQONADGSWGIAPNYPGDVSTSAEAYLALRIIGMSPDSPELFQ  
ARTFIRAAGGLSKMRMFTRIFFAEFGLVPWTAIPQLPAEFILVPAHFPIISYRLASWARSNVVPLLI  
LAHHRPLYPLNGLHKQNPFLDELWLDPATKPLPYGSLDPTDPLSFVFTILDKALSYLGGLRRCR  
TRGYARRRCIQWILQHQEKAGDWAGIIPPMHAGIKALWLEGYKLDHEPIQLGLAAIERFTWTDN  
RGKRLQCCISPVWDTVLMIRALQDTPASLGIKSDPRIADALAWTAENQHRGPEGDWRVYQPNII  
PVGWGAWEYSNTWYPDIDDTAAAVLAFALTHDPATARSRLVRDAVLWIVGMQADGGWAADFH  
ENNRFLFNKI PFSDMESLCDPSTPDVTGRTIECLGMLRDLMLPAEKAGKKEKYGPDGERD  
AADSHLLKIINTACARAIPLYLRTQEATGAWYGRWAVNYVYGTCLVLCGLQYFKHDPFPAEID  
TMATRAVKWLRQIQNSDGGWGESVLSYREPWAGCGPSTPSQTAWALMGLLTVCGGEDRS  
VQRGVRHLVDTQDDILSKGEGAAAWTEREPTSTGFPNHFYISYTLRYVYFPI TALGRYLSLVE  
GKKENGGGA

&gt;seq\_ID 178

MNSINATAAPIDDNVLGDRIGAAATRGLLSLKQSDGHFVFELEADATIPSEYILMRHYLGEVDTV  
LEAKIAAYLRRIQGAHGGWPLVHDGPPDMSASVKAYFALKMAGDSIDAPHMARAREAILSRGG  
AANVNVFTRFLLSFFGELTWRSPVLPVVEIMLLPMWSPFHLNKVSYWARTTMVPLMLVLAALPK  
RARNPGRGIGIRELFLDPATVGTPKRAPHQSPGWALFTGDFRVLRLIEPLSPKWLRRARAKKA  
IAFVEERLNGEDGLGAI FPPMVTVMYDALGFPPEHPRAVTRRGIDKLLVGENEAYCQPC  
VSPWDTALSCHALLEAGGPEAVNSAGKCLDWLLKQELVLKGDWAVKRPDVRPGGWAFOYA  
NGHYPDLDLDTAVVVMAMDRVRRNGPNRGYDEAIARGREWI EGMQSRDGGFAAFDADNLEY  
LNNIPFSDHAALLDPPTEDVTARCVSMLAQLGETVDSSSMAAGVEYLRRTQLAEGSWYGRW  
GLNYIYGTWSVLCALNAGVDHQDPVIRAVNWLVS IQNADGGWGEDAVSYRLDYKGFEGAP  
TTASQTAWALLALMAAGEVENPAVARGIKYLIDTQTKKGLWDEQRYTATGPRVRYLRHYGYS  
KFPPLWALARYRNLSTNSKAVGVGM

&gt;seq\_ID 177

MNATVAQIGDAVLEDRI GSATRGLLNLKQSDGHFVFELEADATIPSEYILRRHYLGEVDTVLEA  
KIAAYLRRIQGAHGGWPLVHDGPPDMSASVKAYFALKMIGDSV DAPHMARAREAILSRGGAN  
NVFTRFLLSFFGELTWRSPVLPVVEIMLLPMWSPFHLNKISYWARTTMVPLMLVLAALPKRARN  
PRDVGIRELFLDPATVTRPKRAPHQSPAWPALFSSLDWILRRIEPLPPKRLRARAMEKAI AFVE  
ERLNGEDGLGAI FPPMVTVMYDALGFPPEHPRAVTRRGIDKLLVIGED EAYCQPCVSPW  
DTALSCHALLEAGAPEALNAGKCLDWLLKQELVLKGDWAAKRPDVRPGGWAFOYANGHY  
PDLDDTAVVVMAMDRVRRNGRDKYDEAI ERGREWI EGMQSRDGGFAAFDADNLEYLNNIP  
FSDHAALLDPPTEDVTARCVSMLAQLGATVDGSSMAAGVEYLRRTQLAEGSWYGRWGLNY  
YGTWSVLCALNAGVDHQDPAIRKAVDWLLS IQNEDGGWGEDAVSYRLDYKGFEGAPTASQ  
TAWALLALMAAGEVENPAVTRGIKYLIDTQTKKGLWDEQRYTATGPRVRYLRHYGYSKFPPL  
WALARYRNLSTNSKVVGVGM

&gt;seq\_ID 170

MREAVSKVEALQORSKTQGISLEDVERGVAQATRALTAHDDGHI CFELEADATIPSEYILFHHF  
RGTQVPDGLLEAKIGNYLRRTQGRHGGWALVHEGPFDMSC TVKAYFALKMIGDIEAPHMRA  
REGILSRGGAANANVTRFMLALYGEVPPWRAVVPVVEVMFLPKWFPFHLDKISYWAR TTVVP  
LFVLQATKPRARNPRGISVQELFVTPPESEVRSWPGSPHATWFWPTPIFGIDRVLQRVENHLP  
KSRQRAMEMARAWVSERLNGEDGLGAI FPPMVTVMYDALGFPPEHPRAVTRRGIDKLLVIGED  
EAYCQPCVSPWDTALASHALLEAGGPEAEQAARAGLDWLPKRVLDIVGDWAARKP  
KVRPGGWAFOYANAHYDLDLDTAVVVMAMDRAMHQGLVAGMPDYKASIRAREWVEGLQ  
SEDDGWAFAFDADNLMHMYLNHI PFSDHGALLDPTADVTARVVGMLSQLGETRETSRALDRGV  
NYLLNDQBEDGWSYGRWGMNF IYGTWSVLCALNAGVD PADPRIQKAVSWLIRIQNPDDGGW  
GEDASSYKIDPAFEPGSS TASQTAWALLALMAAGAVDDPAVTRGINFLTRTQGGADGFWKEERY  
TATGPRVRYLRHYGYPKFPPLWAMARFRNLKRGNSRRVQFGM

&gt;seq\_ID 14

MLLAEVQKALRLAVGHSLDLQRADGAWCGEVHSNATFTSQYVFLQOQIGLPLDPTIEGLSRW  
LFSQQNEDGSGWGLGPGGGDVSTTTETYLALKILGVSPEDPRMAAARTSIIKAGSLPATRMFTR  
VFLASFGILPWSAVPPLPAELILLPTLFPVNIYNLSSWARATCVPLLLIRHHEPLHSLPNGRHAEN  
DFLDELWTKDIPRDFCYTTPLSRMWRGLDYAGIFFTSADHGLRFLGQYFHSPLRNLRRKIIINWI  
LDHQEQSGEWAGYWPQHNNI WALSLEGYSLDHPVLRRIAAVKS FVLHDTGMRAQVTVSQ  
VWDTALMSIALSDSAPSTGI SPTQAI DWLHMHVEVASHRGDWRVLRPKLATGGFCFEFNFTLYP  
DYDDTAAVIMALIKSNPAHLISGCVRRAAQWILGMQNRDGGWGFADWNNDKFFLNKIPFSDMD  
SLCDPSTPDVTGRIIECFGMMAGRHGYSLDCQLENRLRASSQLAIA YLLGCQENNGSWWGR  
WGVNYLYGTSNVLCGLAYYDRSSLKGDVKSNSNI VSAVDRASEWLKARQHSNNGWGE  
ESYDNAQLAGCGQPTASQSAWVTMALNLYLSP TDEVIQRGVSVLVRNQVYKGDSESRATWLE  
RYTATGFPGLHMYEYD YRHYFPIMALGRYVKNKLSGSHKLL

&gt;seq\_ID 180

MTRALRQAPESAGAIGIAAASPATETSQDTHPREISGAI TAARDALLKLOQADGHWCFML  
EADCTIPAEYI LWHTFTGELEPEIERKLAARLRAKQASHGGWPLYEGGDLDISCSVKVYALKLVGD  
DENAPHMRRAREEAILAQGGGARANVTRALALAMFSQIPWRGVFPIVVEIMLLPRWFPFHLKSVS  
YWSRVTMVPALILYSLKQAQONPRNVHIQELFTVPPEQERHYFPVRSRLNKILLSVERTARLLEP  
LTPSMRRLRALKIETWFTERLNGEDGLGGIFPAMVNAHESLILLYGSPDHPWRVQAKKALQNL  
VIEEKNSASCQPLSPIWDTGLAALALQETEGG HTTAPVIRALDWLKERQILEQSGDWQVQHP

-continued

Enzyme Sequences

NLKGGGWAFQYNNSYYPDLDDTALVAWSMDQAATPERYGEAIGRACDWLQCMQSRNGGFA
AFESDNTHYYLNEIPFADHGALLDPPADVTARCIVLLGRLNKPKQYAETLQRALDYLRREQEPN
GSWFGRWGTNYIYGTWSALTALEQANIDPQEGFIRKAVEWLKQVQRLDGGWGEDNYSYFDS
SLAGRYQESTPVHTAWALLALMAVGEANSEAVKKGIAYLQIQEDGLWDHPAFNAPGPPRVF
YLKHYGYDKFFPLWALARYRNHLNRQC

>seq\_ID 155
MMANATDTIELPPSRAADRIVPMTDIDQAVDAHAALGRRQDDGHWFVEADATIPAEYVLL
EHYLRDRIDPALEERIGVYLRRIQGDHGGWPLYHGGKFDVSAITVKAYFALKAI GDDIDAPHMARA
RAAILDHGGAERSNVFTRFQLALFGEVPHWATPVMPELMLLPRKALFSVWNMSYWSRTVIAP
LLVLAALRPRAINPRDVPVPELFTVTPDQVRDWRGPIYRQVSLGRFLKQVYDIALRPAERLIPDTR
QRRAIKAAVDFIEPRLNGEDGLGAIYPAMANTVMMYRALGVSDSPRAATAWEAVRRLVVELDGE
EAYCQPCVSPIDWDTGLAGHAMIEAASGPKGIRPEDTKKKLAAAEEWLRERQILNNGEGLRGDQL
PRRAPRRLGLPVQQRLLPRRGRHSGSRHVLHREGDPADEALERARQWIGMQSSNGGWGA
FDIDNLDLFLNHI PFADHGALLDPPADVTARCISFLAQLGHPEDRPVI ERGIAYLRDQEREGC
WFGRWGTNYIYGTWSVLCAYNAAGVAHDDPSVVRVAVDWRVSVQREDGGWGEDCASYEGAT
PGIYTESLPSQTAWAVLGLMAVGLRDDPAVMRGMAYLTRTQKDDGEWDEEPPYNAVGFPPKVFY
LRYHGYRQFFPLLALSRYRNLASSNSRHVAFGF

>seq\_ID 8
MNRMLQPLHSGAGIFRSSLDRVIAQARQALGGRQAEDGHWCFFFEADCTIPAEYILMQHYMD
ERDEALEARIYVYLRGKQADHGGWPLYYGGHFDLSASVKVYALKLAGDDPELPHMRAREAI
LAHGAERSNVFTRITLALFAQVPRVAVPFIPEIMLLPRWFFPHIYKVASWRTVMVPLFLILCS
LKARAKNPLQVHIRELFRPPDQITDYFSHARRGIVAYIFLSLDRFWRLMEGWIPHGIRRRALKK
AEAWFTARINGEDGLNGIFPAMVNAHEALELGGYPPDHDYRRQTGAALRKLVRERANDAYCQP
CVSPVWDTCLALHALLLEDGEVSPAVQNGIRWLKNRQI GAEPGDWRESRPHLAGGGWAFQY
ANPYYPDLDDTAAVGMALARAGRAEDRDSIEKAAWLAGMQSRNGGFGAYDNDTHYYLNEI
PFADHKALLDPPADVTGRVVAFLAHLARPRDRDLVRRVAVYLLREQESSGAWFGRWGTNYIY
GTWSVLMALAELENDPSLKPMTMERAAYWLRVAVQGDGGWGESNDSYSDPGLAGMGQTSTAA
QTAWACLGLMAAGDRDSVALHRGIAWLQAHQEGDGCWQAPFFNAPGPPKVFYLIYHGYAFYF
PLWALARYRNLCMAHE

>seq\_ID 203
MSMNEAVLAAAPRAAVATAAPALQAPIEALSPLDAGIGHAVDALLAQONADGHWVVEADATIP
AEYVLMVHYLGETPDLSEARLARYLRRIQNADGGWPLFHEGRSDISASVKAYFALKMAGDDP
QAAHMARAREVILAMGGAETSNVFTRITLLALYGVMPWQAVPMPVEMLLPQWPPFHLKSVS
YWARTVIVPLVLVNSLRQARNPRKVGIDELFLGSRDAVRLPAPRAPHQKQHWALFHGADVLL
RTAEHVMPRGLRRRIDLAAKAFVRRERLNGEDGLGAI FPAMANSVMMFDVLPDDPDRARIAR
RSIDKLLVHGD EAYCQPCLSPVWDTALAAHALL EASEPRATAAVTRALDWRPLQVLDVRGD
WTVRRPDRVPGGWAFOYANPHYD VDDTAVVVAAMHRAARTDHSGRADPNAEATARAIEWI
VGMQSANGGWGAFEPENHLYLNNIPFADHGALLDPPADVARSARCLSMCQTGATPKSEPA
ARALQYLLAEQLPDGSGWFGRWGTNYIYGTWSALCALNAAGLGPDAPPLRRAAEWLVAIQNPD
GGWGEDGDSYKLEYRGETAPSVASQTAWALLALMAAGQAHPAVTRGIDYLLRTQQADGL
WHEPRFTAVGPPRVFYLRYHYGARYFPLWALARYRNLEERSGNRQVAVGL

>seq\_ID 165
MREAAVSKVETLQRPKTRDVS LDDVERGVQSATRALTEMTQADGHICFELEADATIPSEYILFH
QFRGTEPRPGLEAKIGNYLRRITQSKVHGGWALVHDGPFDMASASVKAYFALKMIGDDIEAPHM
RAVRAKAILQRGGAANANVFTRILLALYGEVPAWAVPVMPEVVMHLPKWPPFLDKVSYWARCT
MVPFLVIOAKKPRAKNPRGVGVAELFVTPDPSVTRWPGSPHATWPWPIFGGIDRVLQKQTDH
FPKVPQRRAIDKAVAWVSERLNGEDGLGAI FPAMVNSVLMYEVVLGYPPPEHPQVKIALEAI EKL
AEKEDEAVYQPCLSPVWDTALNSHAMLEAGGHQAEANARAGLDLWLPQILDIKGDWAETKP
NVRPGGWAFOYANPHYD LDDTAVVVMAMDRAQRQHLVSGMPDYSESIARAREWVEGLY
SADGGWAAFDADNNHHYLNHI PFDHGLLDDPPADVTARVVSMLSQLGETRATSRALDRGV
TYLLNDQEKDGSWYGRWGMNFIYGTWSVLCALNTAGVDPQSPSEIRKAVAWLIRIQNPDGGWG
EDASSYKLNPEFEPGYSTASQTAWALLALMAAGEVDDPAVARGVNYLVRTQGGDLWSEER
YTATGPPRVFYLRYHYGPKFFPLWAMARFRNLKRGNRSRQVQFGM

>seq\_ID 181
MSISPTFSGSSLQKSSLDHSTISEPFTVVDVNVNGISAVALDDAITRARSALLAQREDGHWCF
SLEADCTIPAEYILMMHFMEIDTALERRIANFLRNQVTDHGGWPLYYGGDFDMSCSVKVY
YALKLAGDSPEAAHVRARNAILERGGAARSNVFTRILLAMRYQIPWRGVPPVPAEIMLLPRW
FPFHLSKVAYWSRTVMVPLSILCTLKAKAANPRNIHVRELFVDPMEKNYFPVTRPLNHLLLYL
ERLGSKLEPLIPSPTRRRALKAEQWTERLNGRDGLGAI FPAMVNA YEALTLGLYDHDHPLLQ
CRLALRELLVNEGEDI TWQPCVSPVWDTVLSLALQEDERADNGPVRHALDVLVPLQALDQ
PGDWRNSRFDLPGGGWAFOYANPHYD LDDTAAAWALCQADTEDYRTSITRAADWLAGM
QSSNGGFAAFDIDNVHYLNEIPFADHGALLDPPSSDV TARCIGLLALNGBARHQETVKRGLTF
LFNEQEPGSAWFGRWGTNYVYGTWSVLEALKLARVDHHDQAVKRAVQWLKSVQRADGGWG
ETNDSYLDSELAGQLETSTSPQTAWAVLGLMAAGEVGS TAVRNGIDYLRITQSAAGLWEEPPWF
TAPGPPKVFYLYKHYGSKYFPLWALNRVYRANMSR.SVV

>seq\_ID 110
MILFPAGFYFSIYEISYWSRCIVVPLSIAIARKPHVTVGDDLLKELYLVPRD VVYRIERDQDGGFC
WYNFFIDADSIFRRYEQHPKIFIRRIAKKMAEKWLLHEHMEKSGGLGAIWPAMINSIFAMKCLDYP
DDHPALTAQMKEVEALVIYEGDMLYLQPCVSPVWDTAWSI IAMNDSGIPGSHVPLQKAGKWLL
SKEVRDFGDWKLKCVKEBPGWYFQYANFYPD TDDTGAVLMLAQRVSLPEDMHKEKTL LRA
LRWLQAMQCDGGGWAFAFRNKNKTI LNNIPFADFNALLDPSTSDVTGRCIEFFGRIGFNKTYL

-continued

## Enzyme Sequences

NIKKAVEFLKKEQDEDEGWSWFRWGSNYIYGTWSVISGLIYAVGEDINKAYIKKAIWLKLSVQMSD  
GGWGETIKSYEDSALKGIGKSTPSQTAWALLTLITAGEIKSSSTERGIDFLLLSTQKEDGWSWDER  
EFTATGFPKVFYLYKHYMYRNYFPLMALGRYRHFTHKLATSQ

&gt;seq\_ID 182

MSISQAFRTLIQKSSLSLVSSENFADDVAGNEANEISAVTLDEAITRAYTALLAQQREDGH  
WCFPLEADCTIPAEYILMMHFMDEVDTVLERKIANFLRTRQVTDGGHGWPLYGGDFDMSCS  
VKTYIALKLAGDSPAAHMHVHARNAILERGAARSNVFRLLAMRYQIPWRGVFPVPAEIMLL  
PRWFFPHLSKVAVYWSRTVMVPLSILCTLKAKAINPRNVHVQELFVVDVVKKEKNYFPVRTSLNRL  
LLYVERLASKLEPPFIPFIRRRRAVKAEQWVIERLNGNDGLGAIIPAMVNAYEALTLGHDRDHP  
LLQQCRQSLRELLVDEGEEITWCQPCVSPVWDTVLATLALQEDKQADSEPIRRALDWIVPLQIL  
DEPGDWRDSRPNLLGGGWAFQYANPHYDLDDTAVAWALIQTGAEDYRVSITRAADWLAG  
MQSSNGGFAAPDIDNAYYLLNEIPFADHGALLDPPSTDVARSVCGLLALNGEVRHQEAVKRGL  
DFLFNEQESSGAWFGRWGSNYIYGTWSVLEAFRLARVDKGHQAVQRAIQWLESVQRADGGW  
GETNDSYLDPQLAGQLEASTSFQTAWAVLGLMAAGEVENTAVRKGIDYLLRQTQIATGLWEEFPW  
FTAPGFPVRYLYKHGYSKYFPLWALNRYRTLSSKSAV

&gt;seq\_ID 162

MSPFLQASDDNNPLFKESCQALDHATEFARDTLVNKEHWCGWVLSNVTVTAEWIFLQYILGLE  
MSNEDRRRGLKHFTHSSQRPDGWSLQTTGGELSCITEAYLALKILGVSPEEDYMRVARDY  
VRSHGGAEKMRMLSRPHLAMPGLIPWAAVPMQMPPELIFMPSWSLVNIYKFSWARCNIVGLCM  
LRVHEPLYALPNGKQLDNDYLDLWLDPYHKAIPYTPVYQLMQTSPGLVLFQLGDLFLWLLS  
LGFWFLRRWAVSSIQWLDHQEPSPDGGIYPPMHNIILALMLEGWSQDDPVIQRGIGACQ  
RFLAEDPAHGKWMQPSVSPVWDTFLMIRAVADAKTTDDADKLLVKPVDVWLAQQIDDDHIGD  
WRIYRPDIPAGGFAFEYFNKWPVDDTAVGVVALMRHDPVSLVNDRLKAAAWTLGMQNRD  
FGWAAFADANNAFYLHATPFSMDSLDSSTPDVTGHVLEMLGLMYRLERQGRVKSPEMLAF  
LSQSHGACDRGLGYSLLGSQEAFFGGWYGRWGVNYIFGTSAAALCALAYFADRKGVRGKMAAGA  
DWLRSRQNPDDGWSGELLESDYDNKALAGRGRSTPSQTAWALQGLELEDPRGEVVEAGVNW  
LLRHQVTSPSRNSGRVSAWPEDDYTATGFPGHFYLKYLELYCHYFPMMLALARYRSCIQDGA

&gt;seq\_ID 172

MDDRVAATFEAQPRAGFSGVEAAISRAREALLAVQKPDGHVFELEADVIPAELYLFRHFLG  
DPAKTEIERKIGVYLRRRQTAAGGWPLFAEGVFNVS SVKAYFALKIIGDDPNAPHMAKARNAIL  
AHGGAAQSNVFRSLLALYGEVPRAVPAMPVEIMHLPRWFFPHLSKVSYWGRVTIAPLIVVH  
ALKPRAKNPRKISVSELEFPVAPAEVSRWPGAPHKSPWTTIFGAIDRVLHKTEPLLPARSHQTAI  
DKAVAFVTVARLNGEDGLGAIYPAMAYSAMMFFALGAPLSDPRIVQIRKAIIDRLLVVKDGEAYCQP  
CVSPVWDTALASHALMESAGORPEARTAPAAAFAEALDWLKLQVLDVKGDWATQNPDPVR  
PGGWAFQYANPHYDLDDTAVVVLAMDRAVKTSPLIAGEEETAYVEAISREWEILLGQSANG  
GFGAPDADNDRDYLYNIIPFADHGALLDPPADVTARCVSMLGQLGERPETSALARAIIDYLLSE  
QEEEGSWFGRWGMNYIYGTWSVLSAFNAVERPADCAATRKAANAALKRINQPDGGWGEDGE  
SYALGYKGNPAPSTASQTAWALLALMAAGEVDAPEVALGLDYLVTQADDGFWDEARFTAT  
GPPRVFYLRYHYGAKFFPLWAMARYRNLKSGNRLKTQFGM

&gt;seq\_ID 24

MLGATREPPIDVQIALHSRDDNQTGLVLRGTRRTVDRVLKGLCSSPCFFCSVSLTMTATLTTMA  
TTATMATTEASKPLEAQARTALTKATNYAWEIFSNRHWCGELESNTVTCHEIFFLYLVYQHID  
PGEQSQRQWLLSQNSDGSWGIAPNYPGDI STSAEAYLALRIIGMSTDSPELYRARTFIRAAG  
GLSKMRMFTRIFFAEFGLVPWTAIPQLPAEFILVPAHFPISTYRLASWARSNVVPLLI IAHRPLYP  
LPLNGLHKQNPFLDELWLDPATKPLPYGSSDPTDPVAFVFTILDKALSYLGLRRSPTRGYARR  
CVQWILLQHQEKAGDWAGIIPPMHAGIKALLLEGYKLHDEPIQLGLAAIERFTWADNRGRKLCQC  
ISPVWDTVLMIRALQDTPASLGIKLDPRIADALAWTAENQHRGPEGDWVYKPNIPVGGWAFE  
YHNTWYPDIDDTAAAVLAFTHDPATARSRLVDRVAVLWIVGMQNDAGGWAADFHENNQFLFN  
KIPFSDMESLSDPSTPDVTGRTIECLGMLRDLMLRPAENAENGEKYGPDGEGDAADAAHLQ  
IINTACARAIPLYLRSQEATGTWYGRWAVNYVYGTCLVLCGLQYFKHDPKFAPEIQAMAARAVK  
WLKQVQNSDGGWGESLISYREFWRAGCGPSTPSQTAWALMGILTVCGGEDRSVQRGVRHL  
VDTQDDTLSSQDGGAAAWTEREFTIREPLHEASQRIGSD

&gt;seq\_ID 26

MATLTTMTATTATTEASKPLEAQARTALTKATNYAWEIFSNRHWCGELESNTVTCHEIFF  
LYVLYQHIDPGEQSQRQWLLSQNSDGSWGIAPNYPGDI STSAEAYLALRIIGMSTDSPELYR  
ARTFIRAAGGLSKMRMFTRIFFAEFGLVPWTAIPQLPAEFILVPAHFPISTYRLASWARSNVVPLLI  
IAHRPLYPPLNGLHKQNPFLDELWLDPATKPLPYGSSDPTDPVAFVFTILDKALSYLGLRRSPTR  
PTRYARRRCVQWILLQHQEKAGDWAGIIPPMHAGIKALLLEGYKLHDEPIQLGLAAIERFTWAD  
NRGKRLQCCIISPVWDTRYVYKPNIPVGGWAFYHNTWYPDIDDTAAAVLAFTHDPATARSRLV  
RDAVWLWIVGMQNDAGGWAADFHENNQFLNKIIPFSDMESLSDPSTPDVTGRTIECLGMLRDL  
MRPAENAENGEKYGPDGEGDAADAAHLQIINTACARAIPLYLRSQEATGTWYGRWAVNYVY  
GTCLVLCGLQYFKHDPKFAPEIQAMAARAVKWLKQVQNSDGGWGESLISYREFWRAGCGPS  
TPSQTAWALMGILTVCGGEDRSVQRGVRHLVDTQDDTLSSQDGGAAAWTEREFTSTGFPNH  
FYISYTLRYVYFPI TALGRYLSLEGGQEKKKKGGGT

&gt;seq\_ID 171

MGKVELHRTSTQDITLDDVERRVTLASKALMRLANADGHWCFELEADATIPSEYILYHHFRGS I  
PTAELEGKIAAYLRRTQSAHQDGWALIHGPPDMSATVKAYFALKMVGDPIDAPHMRRARDAIL  
RRGGAHANVTRINMLALYGEVPRVAVPVMVEVMLLPRWFFPHLQVYVARTVMVPLFVL  
QAKKPRARNRPGIGIRELFEAPERVKRWPAGPQESSPWRPVFAIDKVLQKVEGFFPAGSRA  
RAIDKAVAFVGERLNGEDGLGAIIPAMVNTVLMFALGYDDHPFAVARSVEKLVTVKBEHA  
YVQPCLSPVWDTALAAHALMEAGGTEAERHAKRAMDWLKLQVLDIKGDWAASKPDVVRP

-continued

Enzyme Sequences

WAFQYANPHYDLDDTAVVVMAMDRVQSRSPGPDAAAYGLSIARAREWEVGLQSRDGGW
AAFADANTYHYLNYIPFSDHGALLDPTADVTARCVSMLSQGETRETCPPDRGVAYLLADQ
EADGSWYGRWGMNYIYGTWSVLCALNAAGIDPACFPVRRAVTWTALAIQNPDGGWGEDASSY
KLEYRGERAPSTASQTAWALLALMAAGEADNPVARGINYLTRTQGDGLWAEDRYTATGF
PRVFYLRHYGAKFPPLWALARYRNLQRGNLSLKVAVGM

>seq\_ID 173
MLREATAISNLEPPLTASYVESPLDAAIRQAKDRLLSLQHLEGYVWFELEADCTIPAEYILMMHF
MDEIDAAALQAKIANYLRRHQSADGSYPLFRGGAGDISCTVKVYYALKLAGDSIDAPHMKKARE
WILAQQGAARSNVFTRIMLAMFEQIPWRGIPPTPVEIMLLPKWFPFHLDKVSYWVRTVMVPLFIL
CSHKVTARNPSRIHVRELFVTEPQKERHYFDHVKTPLGKAILALERFGRMLEPLIPKAVRKKATQ
KAFDWFRTARLNGVDGLGAI FPMAMVNAEALDFLGVPDDERRRLARESIDRLLVFQGDVSYCQ
PCVSPIDWTALSTLTLQEVARHTADLRDAALSKGLKWLASKQIDKADPAGDWRVNRAGLEGGG
WAFQPGNDYYPDVEDSAVVAHALLGSEDPSPDDNLRRAANWIAGMQSRNNGGFGAFDADNTY
YYLNSIPFADHGALLDPTADVSARCAMFLARWVNRQPELRPVLERTIDYLRREQEADGSWFG
RWGTNYIYGPQAVLLAYEGRRVNDPDPVRRRAVAWLKSIQREDDGGWGEDNFSYHDPSSYRGR
FHTSTAFQTGFALIALMAAGEXGSPVEVQAGVDYLLRQQRPDGFWNDECFTAPGFPFRVYLYKY
HGVDKFFPLWALARYRNERIALA

>seq\_ID 117
MNETAFANPAPQVGAQRQPAAPQEAARLPAPALDRGIDRALDALLHQQRPDGHVWYELE
ADATIPAEYVLMVHYLGEAPDLLEEARLARYLRRIQNPDGGWPLFHQGRSDISASVKAYFALKM
AGDDPQSAFMQRARQAIHAMGGAEATNVFTRTLLEALYGVLPWKAVPMPMVEIMLLPRWFPFH
LSKVSYWARTVIVPLLVNLSLRPQARNPRGVGINELFVGNCHTVGLPPRAAHQHAGWYTVFRG
LDALLRLAEPLFRTRRRRAIAAQAQFVRERLNGEDGLGAIFPAMANSVMMFVLDVGVPPEDPAR
AVARRSIBERLLVEHGDDEAYCQPCLSPVWDTALAHALLEETGEARAAQAAGRALDWRPLQVLD
LRGDWAVRRPLVRPGGWAQYANAYYPDVEDTAVVAAAMDRFMRAHAPGRYGEAVARAT
EWIVGMQSGNGGWFGEFEPENTHLYLNNIPFADHGALLDPTADVSARCLSMCQTGATPANS
EPAARALRYLLAEQMPDGSWFGRWGTNYIYGTWSALCALNAAGLPEAPELCRAVAWLARIQ
NADGGWGEDGSSYRLDYSYEPAPSVASQTAWALLALMAAGAAQHPAVARGIDYLLRTOQP
GGLWHEPRFTAVGFPFRVYLYRHYGARYFPLWALARYRNLQRGLGDHGGNSQVAVWGL

>seq\_ID 204
MSMNETAFATAVPRIPASAGDSPAPRDAAQALDQIGRAIDALLHQQRPDGHVWYELEADAT
IPAEYVLMVHYLGEAPDLLEEARLARYLRRIQNPDGGWPLFHEGRSDVSVKAYFALKMAGD
DPQAAMQORARRAVHALGGAEASNVFTRTLLEALYGVMPWLAVPMPMVEIMLLPQWFPFHLSK
VSYWARTVIVPLLVNLSLRPQARNPRGVGINELFVGNCHTVGLPPRAAHQHAGWYTVFRGLDA
LLRVAEPLVPTLRRRAIAAQAQFVRERLNGEDGLGAI FPMAMANSVMMFVLDVGVPPDDPARAL
ARQSVRELLVEHGDDEAYCQPCLSPVWDTALAAHALLEETGEARATAAAGRGLDWRPLQVLDV
RGDWAVRRPLVRPGGWAQYANAYYPDVEDTAVVAAAMNRYMRAHDVPGRYDEAVARAAE
WIVGMQSGDGGWGFGEFEPENTHLYLNNIPFADHGALLDPTADVSARCLSMCQIGATPGKSE
PAARALRYLLAEQMPDGSWFGRWGTNYIYGTWSALCALNATGLAPEPEMRRVAWLQIQN
ADGGWGEDGSSYRLDYSYEPAPSVASQTAWALLALMAAGAAQHAAVARGIDYLLRTOQSG
GLWHEPRFTAVGFPFRVYLYRHYGARYFPLWALARYRNLQRGGAHQVWGL

>seq\_ID 79
MRIGTTNPSMFPPLSSGAVFYREVNELREVQOEINRIQAFLLQRQEDGTWRFCLESSPMT
DSHMIILLRTLGIDHERLMEKLTAHITALQHDNGAWKLYPDEQEGHLSSTIDSYYALLSGKYTK
NEPRMALARSFLEKGGTLQANMLTKFATALTGQYQWPSHPLVVEIALLPPSPFVSYFDVFGY
ARVHLAPMMIVADRYNYVKKPDNAPDLSDLYADTPISRGLYPHRFLENFLKEGQSFATIHDSLQ
QLPFLPGQLHLKALRRLRLEQYILARI EPDGTLYNSTSTFFMIFALLARGFSPKDPLIQKAMQGLTG
SVYDYENGAHLQLATSAVWDTALLTFSLQKSGLSPTHAPAIQKANRYLLRKKQHTYGDWIKIRNP
NGKPGGWFSDYNTMNPDI DDTAALRSLRLLAR TDVTAATAWKRGLEWLLSQNDGGWP
AFERNTDADFIRHLP IEGADTVSTDPSSADLTGR TLEFLGNYAGRTL TDHVEKGVRRLLKHQE
SDGSWYGRWGIAYL YGTWAAI TGLMAVGFSPTEPAIQKAVAWLVANQNPDGGWGESQSDL
KKTYYVPLGASTPSTAWAIDALI AVSSKPTABLQRGIRYLLTHNQANDWTTTRYPTGGRRPGGT
YFAYHSYRWIWPALLLSHYQVKYANT

>seq\_ID 70
MLLYDKVHEEIERRTALQTMQRQDGTWQCFEGALLTDCHMI FLLKLLGRNDEIEPFVKRLVS
LQTNEGTWKLYEDEKGNLSATI QAYAAALLASEKYSKEDMNMRAEMP I KEHGGVSRAPFMT
KFLLAIHGEYEFEPALFHFPTPILFLQDDSPLSIFGLSSSAR IHLI PMMI CMNKRFRVEKKLLPNLNHI
AGGGGQWPREERSPLIQSFLGDVKKVI SYPLSLHKKGYEEVERFMKERIDENGLTYSYASATF
YMIYALLALGHSIQSPI I EKAVTGLKSYIWKMDRGSHLQNSPSTVWD TALLSYSLQEAQVTNENK
MIQRATEYLLQKQQTKKVDWSVHASLVAGGWGFSVDVNTTIPIDDDTTAALRALARSRGNDRV
DDAWGRGVEWVKGLQNDGGWGFAPERGVTSLKLSNLP I ENASDMI TDPS TPI TGRVLELFG
TYAPNELLEEQKKAIKWLMVDVQEQNGSWYKQWGI CYI YGTWATMTGLRALGVPSHPALKK
AASWLEHLQHEDEGGWGESQSSVEKFI SLPSTPSQTAWALDALISYDQETPIIRKGISYLLA
QSTMNEKYPTGTGLPGGFYIRYHSYGHYIPLLALAHYVKKYRK

>seq\_ID 140
MAGERSALITALKRSQAADGSRFPFETGISTDAYMI ILLRTLDINDEPLIQALVERIESRQEANG
AWKLFADGEDGNVTATVEAYYALLYSGYRQPTDRHMQKAKRRI LDMGGDRVHLFTKVMLAL
TGQYWPGRFPLPLEFFLLPPSFPLNMVYDLSVYGRANMI PLLIAADSRYSRKTDKSPDLSDLFA
SRGDWGMPESRLLTYVKRSLIGLPAQLHQAAKQRAVRYLFEHIEPDGTLYSYFSSTFLFIFALL
ALGYRNDPDRIRQAVRGLRSLRRTTIDGHVHLQYTTASVWNTALASYTLQEAQVPMTDRAIEKA
NRYLLSRQNVRYGDWAVHNPYSTPGGWGFSVDVNTMNPVDDTTAALRAIRQAAKETAFRH

-continued

## Enzyme Sequences

AWDRANQWLFQMNDGGFAAFEKKNVSRFRWRYLPIEGAEPLLMDDPSADLTGRTLEYFGTF  
 AGLTKDQRAVSRVDWLLSHQERNGSWYGRWGCYIYGTWAAITGLTAVGVPAAHHPALQKAV  
 RWLLSIQNDGGWGESCKSDGAKTYVPLGSDTPVHTAWALDALVAAAERPTLEMKAGFRALF  
 RLLHHPDWTASYPVGGMAGAFYIHYHSYRYIFPLLALAHYEQKFGPLDD

&gt;seq\_ID 137

MAGERSALITALKRSQAADGSWRFPFETGISTDAYMIILLRDLINDEPLIQALVERIESRQEANG  
 AWKLFADDEGDNVTATVEAYYALLYSGYRQPTDRHMOKAKRRI LDMGGLDRVHLFTKVMLAL  
 TGQYWPWGRFPLPLEFFLLPPSPFLNMYDLSVYGRANMIPLLIADSRYSRKTDKSPDLSDLFA  
 SRGDWMPESRSLTYVKRSLIGLPAQLHQAAKQRAVRYLFEHIEPDGTLYSYFSSTFLFIFALL  
 ALGYRNDPDRIRQAVRGLRSLRTTIDGHVHLQYTTASVWNTALASYTLQEAQVPMTDRAIEKA  
 NRYLLSRQNVRYGDWAVHNYPSTPGGWGFSVNTMNPVDDTTAALRAIRQAAKETAFRH  
 AWDRANQWLFQMNDGGFAAFEKKNVSRFRWRYLPIEGAEPLLMDDPSADLTGRTLEYFGTF  
 AGLTKDQRAVSRVDWLLSHQERNGSWYGRWGCYIYGTWAAITGLTAVGVPAAHHPALQKAV  
 RWLLSIQNDGGWGESCKSDGAKTYVPLGSDTPVHTAWALDALVAAAERPTLEMKAGFRALF  
 RLLHHPDWTASYPVGGMAGAFYIHYHSYRYIFPLLALAHYEQKFGPLDD

&gt;seq\_ID 136

MVADERSALIDALKRSQSDVSGSWRFPFETGISTDAYMIILLRDLGIHDEPLIQALVERIESRQDAN  
 GAWKLFADDEGDNVTATVEAYYALLYSGYRQPTDRHMOKAKRRI LDMGGLDRVHLFTKVMLAL  
 TGQHSWPRRFPPLPLVFFLLPPSPFLNMYDLSVYGRANMIPLLVVAERYSRKTDNSPDLSDLA  
 ASRNDWRLPDTEALWSYVKRSLTGLPAWLHRAAEQRAVRYMLEHIEPDGTLYSYFSSTFLLIFAL  
 LLALGYPRDHPHARAVRGLRSLRTEIDGHVHLQYTTASVWNTALASYTLQEAQVPMTDRTIEK  
 ANRYLLSRQHIRYGDWAVHNYPSTPGGWGFSVNTMNPVDDTTAALRAIRRAAKETAFRH  
 AWDRANQWLFQMNDGGFAAFEKKNVSRFRWRYLPIEGAEPLLMDDPSADLTGRTLEYFGTF  
 AGLTKDQRAVSRVDWLLSHQERNGSWYGRWGCYIYGTWAAITGLTAVGVPAAHHPALQKAV  
 RWLLSIQNDGGWGESCKSDGAKTYVPLGSDTPVHTAWALDALVAAAERPTPEMKAGFRALV  
 RMLHHPDWTASYPVGGMAGAFYIHYHGYRYIFPLLALAHYEQKFGPFVD

&gt;seq\_ID 49

MLLYEKVVEIARRTTALQTMQRQDGTWRFCFEGAPLTDCHMI FLLKLLGRDKEIEPFVKRLAS  
 LQTNBGTWKLVEDEVGGLSATIQSYAALLASEKYTKEDANMKRAEMFNERGGVARAHFMTK  
 FLLAIHGEYEPSPFLHPTPI MFLQNDSPLSIFELSSSARIHLIPMMLCLNKRFRVGGKLLPNLNI  
 AGGGGEWFRDRSPVQTLLESVKKIITYPYLSLHKKGYEEVERFMKERIDENGLTYSYATASFY  
 MIYALLALGHSIQSPIQKAITGIAASYIWKMERGSHLQNSPSTVWDTALLSYALQEAQVPSKAVI  
 QNASAYLLRQKQTKKVDWSVHAPNLPFGGWGFSVNTMIPDIDDTAVLRALARSRGDENVD  
 NAWKRAVNWVKGQNDGGWAFEGKVTSRILANLPIENASDMI TDPSTPDITGRVLEFFGT  
 AQNELPEKQKQSAINWLMNVQEEENGSWYGRWGCYIYGTWAVLTGLRSLGIPSSDPSLKRAL  
 WLEHIQHEDGGWGESCSQSSVEKRFVTLFPSTPSQTAWALDALISYYEKETPIIRKGISYLLSNPY  
 VNEKYPTGTGLPGFYIRYHSYAHYIPLLLTAHYTKKRYK

&gt;seq\_ID 62

MNIVIRISKGVSNLLLDEKAHEEIVRRATALQTMQWQDGTWRFCFEGAPLTDCHTIFLLKLLG  
 RDKEIEPFVVERVASLQTNBGTWKLVEDEVGGLSATIQSYAALLASKKYTKEDANMKRAENFIQ  
 ERGGVARAHFMTKFLLAIHGEYEPSPFLHPTPI MFLQNDSPFSIFELSSSARIHLIPMMLCLNKR  
 FRVGGKLLPNLNIAGGGGEWFRDRSPVQTLLESVKKIISYPLSLHKKGYKIEFERFMKERIDE  
 NGTLYSYATASFYMIYALLALGHSIQSMTIQAIAAGITSYIWKMERGNLQNSPSTVWDTALLSY  
 ALQEAQVSKDNKMIQNTAYLLKKQHTKKADWSVHAQALTPGGWGFSDVNTTIPDIDDTAVL  
 RALARSRNKINIDNAWKGVNWI KGLQNDGGWAFEGKVTSKLLAKLPIENASDMI TDPSTP  
 DITGRVLEFFGTYAQNELPEKQIQRAINWLMNVQEEENGSWYGRWGCYIYGTWAVMTGLRSLG  
 IPSSNPSLTRAASWLEHIQHEDGGWGESCHSSVEKRFVTLFPSTPSQTAWALDALISYDTEPT  
 AIRKGVSYLLSNPYVNERYPGTGLPGAFYIRYHSYAHYIPLLLTAHYTKKRYK

&gt;seq\_ID 59

METLIDPEISRLTQRLLDEQEDGAWRYCFENSLMTDAYMIVLIRSLGIKKERLVQELADRLLSQ  
 QEEKGFWKIYRDEVEGNLSATVEAYFALLWSGAVKEKEDENMKRARDCILSGGLDKVHSMTK  
 FMLAAHQYQWDRFPFVPEVILLPTYPVPSFTDFSAVARVHLAPLPLLLKSERIYRKTSTTPDLS  
 YLLKDQEDFSFFREERSFIEYVTSVGEAIAAPPANLNDLAKKTALNYMLARLEPDGSLYSYFSS  
 SFYMI IALLSQGYSRKPPLVVNAIKALISYQCKGDGYPHIQNSPSTIWDTALISHALQSSGVDSRN  
 AQILKASHYLRYHQHTQKGDWASEAPQTAPGGWGFSESNTPDIDDTAALRALKLDAYTDP  
 VKRMAWNRGVKVALSMQNKDGGWPAFEKKNKNDILSWVPMGDAEDAALDRSCADLTGRTL  
 EFLGNDAGMRENSQVLKGI EWLNNQENDGWSYGRWGCYIYGTWAAITGLRAVGVANTHPA  
 HQSI IKAIKWLYQIQNSDGGWGESCRSDKERKYISLGASTPSQTAWALDALISINDHPTKEIDRGI  
 ESLVRLNNTDWRKEYPTGAGLPGRFYIHYHSYPIYIPLLLALSNYKTKFLEVR

&gt;seq\_ID 51

MVLYGRVCAEIERITITALHTMQQDGAWRFCFEGSPLTDCHMI FLLRLLKEEEEIEPFVARLTSI  
 QTNBGTWKLVEDEGRAGNVSTTIOAYAALLASGMYTKEDVNMKRAEAFIQERGGIARSHFMTKF  
 LLALHGGYEPYPRMFYFPTIIFLPLPDSPLSIFELSSSARIHLIPMIMCNMCRFTVSKTILPNLDHIS  
 GSSKSEWFRDRSSLFETILGEVKKFVITYPLSLHKKGDKEAERFMIERIDRNGTLYSYASATFY  
 MIYALLALGHSIQSPLIQAVAGLRTYKWHMEAGIHLQNSPSTVWDTALLSYALQEAQVNVNESP  
 MIQTATEYIWRQHEKHDWLSLHAPLSPGGWGFSDVNTTIPDIDDTAALRALARSRKRNR  
 RIEEAWKGVNWKGLQNKDGGWAFEGKVTNRFLTHLPLENSGDMMDPSADITGRVLEF  
 FGTYAPNELQDHQKRAITWLMNVQENNGSWYGRWGVSYIYGTWAAITGLRAVGVANTHPA  
 LKAVMWLERIQHRDGGWGESCRSSIEKRFVPLSPSTPSQTAWALDALISYDEETPVIRKGISY  
 LLEHAASHQEYPTGTGLPNGFYIRYHSYIYIPLLLTAHYINKRYK

-continued

Enzyme Sequences

>seq\_ID 32  
 MLLYEKAHEEIVRRATALQTMQRQDGTWRFCEGAPLTDCHMIFLLKLLGRDKEIEPFVERVA  
 SLQTNEGTWKLEHEDEVEGGNLSATIQS YAALLASKKYTKEDANMKRAENFIQERGGVARAHFMT  
 KFLLAIHGEYEPSPFLHLPPTMIFLQNDSPFSIFELSSSARIHLIPMMLCLNKRFRVGGKLLPNLNLN  
 HIAGGGGEWFREDRSPVQTLSDVKQII SYPLSLHHKGYEIERFMKERIDENGLTLYSYATASF  
 YMIYALLALGHSLQSSMIQKAIAGITSYIWKMERGNHLQNSPSTVWDTALLSYALQEAQVSKDN  
 KMIQNATAYLLKQHTKKADWSVHAPALTPGGWGFSDVNTTIPDIDDTAVLRALARSRGNKNI  
 DNAWKKGWNWIKGLQNDGGWGAFAEKGVTSKLLAKLPIENASDMI TDPSTPDI TGRVLEFFGT  
 YAQNELPEKQIQRAINWLMNVQEENGWSYWKWGI CYLYGTWAVMTGLRSLGIPSSNPSLTRA  
 ASWLEHIQHEDGGWGESCHSSVEKRFVTLPPSTPSQTAWALDALISYDTE TPAIRKGVSYLLS  
 NPVYNERYPGTGLPGAFYIRYHSAHIYPLLLT LAHYIKKYRK

>seq\_ID 31  
 MSTIHENVRSRQKKTISLRETQNDGWSWFCFEGPILTNAPLILLLTSLGDNDKELIAELAEGIR  
 AKQRPDGT FANYPDRKGNVTATVQGYAGLLASGLYSRSEAHMIQAERFII SNGGLRNVHPMT  
 KWMLAANGLYPWPAHLPLSLPLVIPPFPPLHFPYQFSYAR IHFVPMAVTLNKRFSLKNPNVSSL  
 DNLDHRHMTKNPFTWLRSDQDENRDLSSLFAHWKRLLQI PAAFHQLGLRTAKTYMLDRIEEDGT  
 LYSYASATIFMVYGLLALGVSRRHSPVLRKALAGTKALLTSCGNIPYLENSTSTVWDTALLNYALM  
 KSGISDNDQMITSAARFLRERQKKVADWAVHNPHAEPPGGWGFNSINNTNPDCCDDTAAVLKA  
 PRKLYPASWREGLSWLLSMQNSDGGGSAFEKVNHNPLVRLLESAEAAAIDPSTSDLTGRVLE  
 HCLGEAGLSSDHPQIEKAVQWLI RHQEEEDGSWYGRWGV CYIYGTWAAALTMKACGVSQNH  
 AVKKAIRWLKSI QNEDGWSGESCSAEKTYVPLSYGTLVQTAWAAEALLOQYKTHHQAVTKG  
 ISFLIENRHYEGAAFSYPTGIGLPKQFYIRYHSAHIYPLLLT LAHYIKKYRK

>seq\_ID 48  
 MLLYEKAHEEIVRRATALQTMQRQDGTWRFCEGAPLTDCHMIFLLKLLGRDKEIEPFVKRLAS  
 LQTNEGTWKLYEDEVGGNLSATIQS YAALLASKKYTKEDANMKRAENFIKERRGGVARAHFMTK  
 FLAIAHGEYEPSPFLHLPPTMIFLQNDSPFSIFELSSSARIHLIPMMLCLNKRFRVGGKLLPNLNLN  
 AGGGGEWFREDRSPVQTLSDVKQII SYPLSLHHKGYEIEVERFMKERIDENGLTLYSYATASFY  
 MIYALLALGHSLQSSLIQKAIAGITSYIWKMERGSHLQNSPSTVWDTALLSYALQEAHVPKDHKM  
 IQOTIT YLLKQHTKKADWSVHALALTPGGWGFSDVNTTIPDIDDTTAVLRALARSRGNENIDN  
 AWKKGWNWIKGLQNDGGWGAFAEKGVTSKLLANLPIENASDMI TDPSTPDI TGRVLEFFGT  
 QNELPKKQKQSAINWLMNVQERNGSWYWKWGI CYIYGTWAVMTGLRSLGIPSSNPSLKRAL  
 WLEHIQHEDGGWGESCSQSSVEKRFVTLPPSTPSQTAWALDALISYD KETPTIRKGVSYLLAN  
 PYNKRYPTGTGLPGAFYIRYHSAHIYPLLLT LAHYTKKYQK

>seq\_ID 34  
 MNIVIRISKGVSNLLLYEKVHEEIVRRATALQSMQRQDGTWRFCEGAPLTDCHMIFLLKLLG  
 RDKEIEPFVKRLASLQTNEGTWKLYEDEVGGNLSATIQS YAALLASEKYTKEDANMKRAEMFIN  
 ERGGVARAHFMTKFLAIAHGEYEPSPFLHLPPTMIFLQNDSPFSIFELSSSARIHLIPMMLCLNKR  
 FRVGGKLLPNLNLNHIAGGGGEWFREDRSPVQTLSDVKQII SYPLSLHHKGYEIEVERFMKERID  
 ENGLTLYSYATASFYMIYALLALGHSLQSSMIQKAIAGITSYIWKMERGNHVNQNSPSTVWDTALL  
 SYALQEAHVLDKNMQLQATAYLLKQHTKKADWSVHAPALTPGGWGFSDVNTTIPDIDDTT  
 AVLRVLRASRGNEKVDHAWQKGINWVKGLQNDGGWGAFAEKGVTSKLLANLPIENASDMI TD  
 PSTPDI TGRVLEFFGTYAQNELPEKQKQSAINWLMNVQEENGWSYWKWGI CYIYGTWAVL TGL  
 RSLGIPSSDPSLKRALWLEHIQHEDGGWGESCSQSSVEKRFVTLPPSTPSQTAWALDALISY  
 DKETS VIRKGISYLLSNPYINETYPTGTGLPGAFYIRYHSAHIYPLLLT LAHYAKKYRK

>seq\_ID 47  
 MLLYEKVHEEIVRRATALQTMQRQDGTWRFCEGAPLTDCHMIFLLKLLGREKEIEPFVERIAS  
 LQTNEGTWKLYEDEVGGNLSATIQS YAALLASKKYTKEDANMKRAENFIKERRGGVARAHFMTK  
 FLAIAHGEYEPSPFLHLPPTMIFLQNDSPFSIFELSSSARIHLIPMMLCLNKRFRVGGKLLPNLNLN  
 AGGGGEWFREDRSPVQTLSDVKQII SYPLSLHHKGYEIEIERFMKERIDENGLTLYSYATASFY  
 MIYALLALGHSPQSSMIQKAIAGLTSYIWKMRGSHLQNSPSTVWDTALLSYALQEARVSKDNK  
 MIQNATAYLLKQHTKKADWSVHAPALTPGGWGFSDVNTTIPDIDDTAVLRALARSRGNKNI  
 DNAWKKGWNWIKGLQNDGGWGAFAEKGVTSKLLANLPIENASDMI TDPSTPDI TGRVLEFFGT  
 YAQNGLEPKKQKQSAINWLMNAQEENGWSYWKWGI CYIYGTWAVMTGLRSLGIPSSNPSLKRAL  
 SWLEYIQHEDGGWGESCHSSVEKRFVTLPPSTPSQTAWALDALISYDTE TPAIRKGVSYLLSN  
 PYNERYPTGTGLPGAFYIRYHSAHIYPLLLT LAHYLKKYRK

>seq\_ID 52  
 MRSILEDVKAFFRQKTLAELQNRQSDGSRWRFCEGVPVMTDSFFILMLTSLGDQDSSLIASLAER  
 IRSRQSEGDGAFRNHPDERAGNLATVQGYTGMLASGLYDRKAPHMQKAEAF IKDAGGLKGVH  
 PMTKWMLAANGLYPWPRAYIPLSPLLI PSYPLHFPYHFSYAR IHFVPMATFNRRFSLKNQIG  
 SLRHLEAMSKNPLFWLNI RAFDERTFYFNLQWKQLPQWPAYVHQLGFEAGKGYMLDRIEE  
 DGTLYSYASATMFMFISLLAMGISKNA PVVKAVSGIKSLISSCGKEGAHLENSTSTVWDTALIS  
 YAMQESGVPEQHSSSTSSAADYLLKRQHVKKADWAVSNPQAVPVGWGF SHINNTNPDLDLDTA  
 AALKAI PQRPPDAMNRLAWLLSMQNKDGGFAAFAEKD VDHPLIRNLPLESAEAAAVDPSTAD  
 LTRVHLHLGLKGRFTDNHPAVRRALRWLDH HQKADGSWYGRWGVCFYIYGTWAAALTMKAC  
 GVSANQTSVKKAI SWLKS IQREDDGWSGESCSCEAKRFVPLHFGTVVQSSWALEALLOQYERP  
 DDPQI IKGIRFLIDEHESSRRELEYPTGIGLPNQFYIRYHSAHIYPLLLT LAHYAKKYRK

>seq\_ID 188  
 MRSELLQLQSDAGSWRLCFDSGTMPDSYFII LRMLGYSQDEALIRQIASRILSRQLPNGTWKIY  
 PDEEDGNLDATAEAYFALLYSGLFTKLDPRMQLAKQFILSKGGLSKIRSLLTQAI FAAAGQASWP  
 KSMRIPLEVFPSDNGIGDLDFSLSGHARVHIVPI IMLANAQFVQHSASMPDLSDFAGSSKRFEN  
 DSPWIAALATLIGLSLSELLPFESPTPQEKAVQFLPDRLEPDGTLTYTTATMFMILVLLMLGYS

-continued

## Enzyme Sequences

SSSPLIHRMVGSIHVSICANSHVQIASSEVVDTAMLVHALRKAGVNPTS TALENAGAYLRQRQQ  
 TQLGDWAI RNP GTPAGGWGFSNVNTLYPDVDDTTAALRAIQPYSRTP ELQADWQRGLNWVL  
 TMRNDNGGWP AFERQGSRLPI TFFNFEGAKDIAVDPSTVDLTSRTLQFLGQELGMNAGNSWIE  
 STLRWVLSQQESNGSWYGRWGITYVHGTSAAALQGLTAVGIAEDHPAVKKGVDWLLQVQNE  
 GGWGESCI SDKVRVYVPLNFS TPSQTAWALDGLTAAALPKPTPALERGV DALLQSLDRHDWY  
 TYPTGGALPGSVYAHYASNNYIWPLLLALSNIWQKYS

&gt;seq\_ID 60

MGTLQEKVRRFQKKTITELRDRQNADGSWTFCEGPI MTNSFFILLTSLDEGENEKELISSLAA  
 GIHAKQQPDGTF INYPDETRGNLTATVQGYVGM LASGCFHRTEPHMKKAEQFI SHGGLRHHVH  
 FMTKMWLAANGLYPWALYLP LSLMALPPTLP IHFYQFSSYARIHFAPMAVTLNQRVFLINRNI S  
 SLHLLDPHMTKNPFTWLRSDAFEERDLTSL LHWKRVFHPAPFAQQGLQTAKTMYLDRIEKD  
 GTLYSYASAT IYMVYSLSLGVSRYSP IIRRAITGI KSLVTCKNGIPYLENSTSTVWD TALISYALQ  
 KNGVTE TDG SVTKAADFLLERQHTKIADWSVKNPNSVPGGWGFSNINTNPNDCDDTTAVLKAI  
 PRNHS PAAWERGVSWLLSMQNNDDGGFSAFEKVNHP LIRLLPLESAEDA AVDPSTADLTGRV  
 LHFLGKVGTFEKHQH IQRVAVKWLFEHQEQNGSWYGRWGV CYI YGTWAALTGMHACGVDRK  
 HPGIQKALRWL KSIQNDGSGWGESCKSAEIKTYVPLHRGTIVQTAWALDALLTYENSEHPSVVK  
 GMQYLTDSSSHSDSLAYPAGI GLPKQFYIRYHSYPYVFSLLAVGKYLDSIEKETANET

&gt;seq\_ID 56

MQDFKTKVNVYMDLHMOMQHRQREDGAFVFCFEGSMMTNAPLIMLLKAVGDTDQALVHQL  
 AEAI REKQNE DGSFSLYHDQAGHV TATVQGYCGMLVSGRYQQDEPHMEKAARYIRSKGGLKD  
 VHFMTKMWLA VNGMHPWPYFYAPLSFLLIPTYPFLHFYHLSAYARIHFVPMMI ALNKRYTSHEQ  
 FPSLSHL DANMSKNPFDWFMAREERSTHFLAYMRSY TALDSRDFPFGYEAAKRYMFDRLK  
 DGTLYSYLSASIFMVYALMSLGYSPGHHLILKAVKGMKQLVTDCCGKKYAE NSTSTVWD TALV  
 SYASQRAGRTQDDPVIKKSFTYLLNRQOMKADWAIHNRHAAPGGFGFSDLNTNPNDCDDTQ  
 IVLKAI PQTYAPVQWKRGF DWLLSMQNRDGGFSAFEKNQDHFLLRHLPLESAEDA AIDPSTPDI  
 TGRV LHLIASEENDKSPLMQRQKDHCVKWL LDHQEKDGSWYGRWGV CYI YGTWAALTGLKA  
 SGI PSSHPAVQKACRFLKTIQLEDGSGFESCKSSEVKRYVPLPFGTVVQTAWAAEALLQYVQP  
 DDKSILKAI SFLIQHQS KALHYVPGI GLPKQFYI TYHSYFPVFPMMACSTFLEEMRRKNE

&gt;seq\_ID 58

MKNRKGAGCMQLVKSEIERLKQQLLEQTPDGSWNHPDFTGCMTDIYIMVLLRTEEEDEEEE  
 LIKELAKGILSRQGDGAWRLFDHHEGSLSLTI EAYYALLYSGYEEKNHPALVKARRVITKGGG  
 LKKAAGMYTKIMLAL TGQYPWLLFPVPM EIVLLPRSFPLNMYDISVFGRSNLI PVILLGNKFKSRK  
 TALS PDLGDL SVRDDDWPPELRSAEWRSLTSFLAAGVKALVGI PRQIRAWSIEKAREYMQSH  
 TEPDGTLYNYPSTFYMIFALLALGGGPEEPAIRNAVAGLKRMTVKADGRTHI QYTTAAVWNTA  
 LISHALQEAGVPKADNLFQKANKYLAGQQHRRFGDWIVHNTKAEPPGGWFSR FNTPNDVDDT  
 TAALRSLYQPAREKPHYDDIWKKGLLWTL SMQNRDGGWPAFERNVDKLLHLLPIQGAEIFILT  
 DPSTADLTGR TLEFLGKAGYADASLPP IKKAVKWLKHQEPNGSWYGRWGI CYI YGTWAAVTG  
 MAAVGVTLEDKSMKGI DWLLSIQNE DGGWGESCRSDMEKKYI PLKESTLTQTAWADALAAA  
 GMADSTPSRKGAAFLVREGKRKDWTADYPMGQGMANFFYIHYHSYRCIWPLLLALSHYIEKSEA  
 PD

&gt;seq\_ID 57

MQDFKTKVNEYIDELHMQLRQREDGAFVFCFEGPMMTNAPLIMLLKAVGDSQALVHQLA  
 EAIREKQNE DGSFSLYHDQAGHV TATVQGYCGMLVSGRYQQDEPHMEKAAHFIRSNGLKQV  
 HFMTKMWLA VNGMHPWPYFYAPLSFLLIPTYPFLHFYHLSAYARIHFVPMMI ALNKRYTSHEQF  
 PSLAHL DANMSKNPFDWFMAREERSTHFLAYMRSY TALDSRDFPFGYEAAKRYMFDRLK  
 DGTLYSYLSASIFMVYALMSLGYSPGHHLILKAVKGMKQLVTDCCGKRKYAE NSTSNVWD TALV  
 YASQQAGRTQDDPVIKKSFTYLLNRQOMKADWAIHNRHAAPGGFGFSDLNTNPNDCDDTQ  
 IVLKAVPQTYAPVQWKRGF DWLLSMQNRDGGFSAFEKNQNHFLRHLPLESAEDA AIDPSTPDI  
 AGRV LHLIALEENSMSPLMQRQKDHCVKWL LDHQEKDGSWYGRWGV CYI YGTWAALTGLKT  
 AGISSSHSAVQKACRFLKTIQLEDGSGFESCKSAEVEKRYVPLPFGTVVQTAWAAEALLQYVQP  
 DDKVILKAI SFLIQHQSSEALHYVPGI GLPKQFYI TYHSYFPVFPMMACSTFLEEMRRKNE

&gt;seq\_ID 61

MGTLQEKVRRFQKKTITELRDRQNADGSWTFCEGPI MTNSFFILLTSLDEGENEKELISSLAA  
 GIHAKQQPDGTF INYPDETRGNLTATVQGYVGM LASGCFHRTEPHMKKAEQFI SHGGLRHHVH  
 FMTKMWLAANGLYPWALYLP LSLMALPPTLP IHFYQFSSYARIHFAPMAVTLNQRVFLINRNI S  
 SLHLLDPHMTKNPFTWLRSDAFEERDLTSL LHWKRVFHPAPFAQQGLQTAKTMYLDRIEKD  
 GTLYSYASAT IYMVYSLSLGVSRYSP IIRRAITGI KSLVTCKNGIPYLENSTSTVWD TALISYALQ  
 KNGVTE TDG SVTKAADFLLERQHTKIADWSVKNPNSVPGGWGFSNINTNPNDCDDTTAVLKAI  
 PRNHS PAAWERGVSWLLSMQNNDDGGFSAFEKVNHP LIRLLPLESAEDA AVDPSTADLTGRV  
 LHFLGKVGTFEKHQH IQRVAVKWLFEHQEQNGSWYGRWGV CYI YGTWAALTGMHACGLTESI  
 PYYKRLCVGSNPNYKMT EAGENPAKAPKSKHMYRFIEEPLYKRPGL

&gt;seq\_ID 50

MAEAI SYPRRVHII TTKFPVNFYDFSVFGRSNIAPILLADSKFQIPKTTETPDISHLYVRELYWWS  
 EDGRWNGPTKAIKGNVNLIGLPNELHTLGRKQAE NYMLDRLEDDGTLSSYSSSTFFMIYALLS  
 VGYTKDHKVI KKAARGLLSMNTTVKDTIHI QYTTAHIWNTSLI SHALQTAGAS PDDTMVMRANH  
 YLLQRQHTKFGDWAIYQPNLGPGGWGF SHSNTFNPDVDDTTASLRSIQNSLHSHPNYQSSWY  
 RGLSFTLGMQNDGGFPAFEKGVDKTFLHLLPVQGA EFLLPDTPSTPDLTGR TLEFLGESAHLY  
 KDSGAIKRGVNWLIENQRDGSWYGRWGI CYI YGTWAALTGLQAVGVSKHEHPSVQEGIDWLK  
 SIQQDDGGWGESCESDSQKTYIPLSKSTVTQTAWAVDALIAYEKEETVEIKKMEYLLLENWNH  
 EDWTMDYPMGQGMKAFYIHYHSYRYVFP LLTMGHYMRKFM

-continued

## Enzyme Sequences

&gt;seq\_ID 199

MSETISCQRIQAAYQSRPAELLSLRNSTGHWTGELSTSAALSTATAIMALEMIRKRLPADLSLNT  
 YIDNGIRWLAEHQNSDGGWGDVTKSFSNISTTMLCHAVFHATKSTEQYVSHVNNARQYIDRVG  
 GVEAVVARYGKDKTFSVPI LTHCALAGLVKWKTI PALPFELACLPAKFYKTVRLPVVSYALPALIA  
 IGQVRHHPCKPRNP ITRLIRKLAVKRSLKLLI SI QPSNNGGFLEAAPLTSFVMTSLAGMGLTDHPV  
 VQKGLQFLDLSVRPDGWSWPIDTNLATWTTLSVNALEGLTAEFEKTPIREWLLQQYKELHPYV  
 SAEPGGWAWTDLPGGVPDADDTPGAILALLNLQDPEPTQQPADLQVALRNGVKWLLDLQNS  
 NGGWPTFCRGGWALPFDQSAAI SAHVIRALQAWLQTEPESEAELRLRAERAVRKCFCYLAT  
 VQRPDGSLPLWFGNQHVENDENPVYGTARVLAAYAQGEQCSIQAEQGI LFLKSVQNLGDG  
 GWGGAT SAPSSVEETALAVDTLLALGLEPADPVVAQGLNWLSGRVENGTYTETTPIGFYFAKL  
 WYFEQLYPI IFTV SALHRAETVLKKSADDNLRSLSEEDYPIMS VKEK

&gt;seq\_ID 75

MDQDRLQRCYAIARDDLQAQRNGQGHWTGELSTSAALSTATAVSAALQLVVRHDPAQSERLMPLI  
 EGGVRYLTHEQNPDGGWGDTRSYSNIAATTMLVAALTAERREALFEQLAFAENYIEAQQGIP  
 GLRRRYGKDKTFVAVPILNTYALAGLVWREVSPLPFELACLPAKFYKLVKLPVVSYAI PALVAI  
 QARYFHRPFPNLMRGLPFGAAVKKS LAVLERMQPASGGYLEAAPLTSFVMSLAI GNASHPV  
 AQNGVQFLVDSVREDSWPIDSNLANWTTLS ISALATGGDDIAELDCLPWVLANQYQETHPF  
 TGADPGGWGTDLSGSPDADDTPGAMLAI AHFFHS PRADNETRRQIASAASGARWLLDLQ  
 NSDGGWPTFCAGWGTQPPDRSGDLTAHAI RALHAWRSELGDL PVERAI ERGLRYLQKQOR  
 DDGSLPLWFGNQD IHDDENPIYGTVKVLLAYRDLGKMSSETAQRGAAWLAARQNEGGGFG  
 GGPSISTLCGGPGESEVETALAI EALFAAENSNI SAETVPPAVGWLCQRVEEGSYVNTCTPIGFY  
 FSKLWYKLYPRVMTVTS LGAALQANASVPPAPETVTTSSDH

&gt;seq\_ID 325

MATSDPSLAEAI QNTRAHLLSLRNARGHWEGLSNSALSTATAIVALHLVDAPLHSAIAQGV  
 WLVLHQNKDDGGWGD TTKSNLSTTLLCWSALS LCEPDRTEPIQHCEAWIKERTGSLEPEVIC  
 RAVVARYGKDKTFSVPI LMLCAI GGRLGPEKEAWSRVLAALPFELAAPREWFAGI GLPVVSYAL  
 PALIAI GYARFVHAPPSLLNPLHALRKAALWPRIS PMLKLLQPS TGGYLEATPLTSFVTMALASAG  
 EKFPVCVPEAVRFL EDSQRPDGWSWPIDTNLATWTTLS TKALATATSEGREALDIPALKSWLLEQ  
 QYQE IHPFTNAAPGGWAWTDLPGGVPDADDTSGALVALWHLCEDEAERQALAPAVAKGVQW  
 LMDLQNRDGGIPTFCRGGWGTQPPDRSGDLTAHAI RALHAWRSELGDL PVERAI ERGLRYLQKQOR  
 ARPPSRGAPGFNHVPLWFGNEHAKEEENHYGTAQIMNHLSSGLNTPEIKVI LETGHRNLLA  
 WQQLDGGWGSSETGPASLEETAVSVAALALH LTHAGNRTRSSAEDAVAKGTQWL VQHTATG  
 TTFPSAPIGLYFARLWYHEQLYPIVITWTLGALHAVETLSAALPLRARASAPQHPGVVTRKPIHI  
 APPSDP

&gt;seq\_ID 135

MIPAERLR TAYRTARAALLAERVPEGHVWVWELSTSAALSTATAVMALHLVNPFFTHRELIDAGRKW  
 LAHQNDADGGWGDVTKSFSNISTMLCRAAFKLAGKEKYPETVQRVEEYLSRNAGALPTARAA  
 AIRARYGKDKTFSVPI LMTCAVAKLVWDEVPRLPFELACLPAQSWYRFAKLPVVSYALPALIAI  
 GQIHHHRRSQNP IRNTVRR LARGLSLKVLRRIQPTSGGYLEATPLTSFVVMALSSIRRRRAAE  
 QQVIDEGVRFLVAVSRPDGWSWPIDTNLATWTTLSVNALEGLTAEFEKTPIREWLLQQYKELHPYV  
 ERHPYTGADPGGWAWTDLPGGVPDADDTPGALIALAHLDPKSDPQAVLSGLRWRVLRQNDG  
 GGAPTFCRGGWGTLPDRSGADLTAHSVRS LASWYRVWVWAGPPPIEHLRRLKDLFPLSGLF  
 WDVARRNPRFVRYLKKQQRSDGSLPLWFGNQHPADDINPVYGTARVLAAYRDLLEKDAPE  
 CRRGIEFLLSVQNDGGWGGAKGCPSSVEETALAVEVLLDLADGDVAVQKGVAWLAEAVESDR  
 FRDASPIGFYFAKLWYFEKLYPI IFTVAALGRAVKITSPAPAAESA

&gt;seq\_ID 115

METLSRSLAEALAKATQALL TELNPAGHWGELSSSALSTATAIVALGAVDREQQRELIAGGM  
 RWLAHQNDADGGWGDVTKSFSNISTTALCWAASVTS TEHAESA KAEAWLTRAAGSMAQLV  
 PAIEARYGKDKTFSVPI LMLHAI CGRVSWSQI PALPFELALPHQLFGALQLPVVSYALPALIAI  
 GQIHHHRRSQNP IRNTVRR LARGLSLKVLRRIQPTSGGYLEATPLTSFVVMALSSIRRRRAAE  
 QQVIDEGVRFLVAVSRPDGWSWPIDTNLATWTTLSVNALEGLTAEFEKTPIREWLLQQYKELHPYV  
 ERHPYTGADPGGWAWTDLPGGVPDADDTPGALLLHLGVVDAPTRQAGQIGVRWLLDLQNRD  
 GIPTFCRGGWALPFDSPDLTAHTLRAWTAWLPQLDESLKRRTRLRAVTKAIHFLATHQRTDG  
 SWLPLWFGNEHAPDENPLYGTAKVVI ALRELLNRDFTLPNGMLERALCWLVERQDISGGWS  
 GAKNGPVSEETALAVEALAGTGHVSA TDRGAAWLTEQIEADTWREPAPIGFYFAKLWYERL  
 YPQIWTVVALGRVAALRVGESESDTPAGLHRATSET

&gt;seq\_ID 208

MMAVVENSVEVLDRRELRLGTLDDLRLGELLAQRTKDGHWGELSSASALSTATAISAMSAVRS  
 GKLGAADKAALEEQIQSGRRWLADQNDGGFGDTRSHSNIA TSYLVLAAWTLDSDQVTGET  
 TDANAI SRLRNWIQLAGELDGLRRRYGKDKTFVVPILTNMAIAGLVWPKKVSALPFEEAVVPQS  
 MYRFVGMPPVSYAVPALVAIGQVKFLEGGGCLPPWSLVVRAAI EP SMKVLRSMPSSGGYLE  
 ATPLTAFVVMLSASGRADHEVTQNGRFLRDSMLPDGWSWPIDTNLANWATSLATTAL TMDPD  
 DDRSWS TNELIQWQRGCQYQERHPFTGADPGGWGWD TLTGSPDADDTPGAI I SLRMQATT  
 RPDPLCDDYSRDWPAASDSSGSVSNALD TWKACDRGVDWLGLQLNRDGGWPTFCRGGWGL  
 PFDSSNDLTAHALRAIACLPKRESAKRSRAVQRGLRFLRKNQADGWSWLPWFGNQDRPEE  
 DNPIYGTSRVLDVSPALGHDAI SRGLYYLINSQNSDGGWGGESVRETFGLPEGFISSVEETA  
 LAVEALVSWWGRIPGNEGGQAAENDIPDGS PWDASMRSAALRAILSGTRWLDIVAQRERHQV  
 AWPIGFYFAKLWYERLYPLVYTTAALGRVMQRDELRL

&gt;seq\_ID 247

MEIQDEVDLLEPQESLTASADS AVDRALFWLLDAQYEDGYWAGILES NACMEAEWLLCFHVLG  
 IANHPMSRGLVQGLLQQRADGSDVYVYGARAGDINTTVEVYALRCQGYAADHPDIKRARD

## Enzyme Sequences

WIQLQGGVKQVRVTRFWLALIGEPWEETPNLPPPEILFFPRWFFPNHYHFAAWARATLVPLCI  
LSARRMVPVPLNKKSLQELFPEEDRSVAVLGGKAGAWSTFFVHADRALKKYQRTFKRPGRQ  
QAIKMCLEWILRRQDADGAWGGIQPPWYISLMALKAEYVTHPVMKGLAALDAHWSYERP  
GGARFVQACESPVWDTLLSSPALLDCGFSCTSSSELKAVDWI LDQQVLLPGDWQKLPVTS  
PGGWAFERANVHYPDVDDTAVALIVLAKVRPDPDTARVNLAI ERGLNWLAFAMQCRNGGWA  
FDKDNDDKLLTKIPFSDGETIDPASVDVTAHVLEALGLLGYRTHPAVAKALEFIRSEQENDGC  
WFGRWGVNYIYGTAAVLPALASLNMNMNQEFIRRAANWILGKQNDGGWGESCASYMDDTQ  
RGRGPSTASQTAWAMMSLLAVDGGTYAESLLRAEAYLKTQTQTEGTWDEPYTGTGPPGYGI  
GRREIKRQRSLQQHAELSRGPMINYNLYRHYFPLMALGRLAALRGA

>seq\_ID 148

MTSPFKHPISHALTSFNIGVTEPEQSVQKAGAKVHQFPASLWKS KPGKAKSPLDIAIEGCRDF  
FFREQLPKGYWMAELESNTITAEYIMLFNLSLVDHERQRKMSNYLLSKQTEEGFWTIYYGG  
PGDLSTTVEAYFALKLTGYPADHPAMVKARAFILEKGGVIKSRVFTKIFLALFGEFDWLVGVPSP  
VELNLLPNWAVVNYEFSSWARATIIPLSIVMLKRPVHKLPSPQRVQELFVRPPRAIDYTFKED  
GIPTWKNFFI GLDHLKVIYERSPVRPFKKRANGKAEWVLEHQEETGDWGGIQPAMLNAVLA  
LSALGYDNGHPAVAHGLKALENFCEI EDEQIVLQSCISPVWDTALALKALVDAGVPSDHPSLVK  
GAQWLLEREVRRPGDWRVKS PDLEPGGWAFEFLLNDWYPDVDDSGFVMIALKGVEVKDRKAM  
NAAVKRGIDWCLGMQSKNGGWGAFDKDNTRHILNKIPFADLEALIDPPTADLTGRMLELMGT  
GYAKTYPAAQRALKFLKENQEPGEPWWRGWVNYLYGTWSVLCGLAAIGEDLEQPIYIKAVN  
WIKSRQNDGGWGETCESYHDPTLAGMGESTASQTGWALLGLMAAGEVHSATVVRGVQYLI  
STQSQDGTWDETYTGTGFPKYFMIKYHIYRNCPLMALGTYRTRLTGGTA

>seq\_ID 149

MTSPFKHPISHALTSFNNGFAEPEQCVQQTGAKVHHLPASIWKRKMGKAKSPLDVAIEGSRD  
FFFQEQLPKGYWMAELESNTITAEYIMLFHFLGLVDRERQRKMSNYLLSKQTEEGFWPIYYG  
GPGDLSTTIEAYFALKLSEGYPADHPALAKARAFILEQGGVVKSRVFTKIFLALFGEFEWQGVPS  
MPELVNLLPDWYINIIYEFSSWARATIIPVLSVVMHSRVRVPPSARVQELFVRQPTAADYSFA  
KNDGIFTWENFFLGLDRVLKVYEKSPLRPFKNMALAKAEWVLEHQEETGDWGGIQPAMLNA  
VLALNVLGYQNDHPAVEGGLRALANFCIETEDQLVLQSCISPVWDTALALKALLDAGVPPDHP  
SLVKGAQWLLDKEVTRPGDWRVKS PALEPGGWAFEFLLNDWYPDVDDSGFVMIALKGIQVKDR  
KSMDAAIKRGINWCLGMQSKNGGWGAFDKDNTRHVLNKIPFADLEALIDPPTADLTGRMLELM  
GTFNYPITLPAQRALFELKKNQEPGEPWWRGWVNYLYGTWSVLCGLAAIGEDMDQPIYIRKA  
VNWIKSRQNDGGWGETCQSYHDRTLAGVGETSPQGTWALLGLLAAAGEMHSATVVRGVQY  
LISTQNSDGTWDETYTGTGFPKYFMIKYHIYRNCPLMALGTYRTRTRTQP

>seq\_ID 216

MTDVLTRLESPNSTRDRVRSVSSARQYLLSLQHEEGWKGELDTNVMTAEADLLLRQFLGIS  
DEQVTQETARWIRS QREDGTWATFHGGPPDLSTTVEAYVALRLAGDAMDAAHLKAREYIL  
DSGGIESTRVFTRINLALFGEWPSRLPVLPEMMLLPDWPLNIYDASWARQTVVPLTIVG  
SLRPRTRDLGFSVRELRTGIQRDLSPSWAGVPHGLDSVLRLEKLPKLRKVALARAEQWI  
LDRQESDGGWGGIQPPWYISIALHLRGLYPLDHPVLRKALDGLDGTIRHRTENGWIRKLEAC  
QSPVWDTALAMTALDLSGTPPNDPALVRAADWILRQEI RVSGDWRVRRPALEPSGWAFEFAN  
DHYPTDDTAEVVLGLQVRHPEPHRVNAAVERA TAWLVGMQSDGGWGFADNTRTLCE  
KLPFCDFGAVIDPPSADVTAHIVEMLAARGMADSESARRGVRLLEHQEVDGWSWFRGWAN  
HVYGTGAVVPALVACGISPQHEAVRAAVQWLVAHQNDGGWGEDLRSYVDRTWVGRGTSTP  
SQTAWALLALLAAGERGEVVRGVWELMAAQRPDGGWDEPQYTGTFPGDFYISYHMYRIV  
FPLTALGRYLGRGGVDVGTG

>seq\_ID 229

MTATTDGSGTASLRPLAASADTDITIPAAAAGVPEAAARTRRATDFLLAKQDAEGWKGDL  
ETNVMTDAEDLLLRQFLGIDEETTRAAALFIRGEQREDGTWATFYGGPGLSTTIEAYVALRL  
AGDSPEAPHMARAAEWIRSRGGIASARVFTRIWLALFGWKKWDDLPELPELIYFPTWVPLNI  
YDFGCWARQTI VPLTIVSAKRVPVPAPFPLDELHTD PARPNPPRLAPVASWDGAFQRIDKALH  
AYRKVAPRRRRAAMNSAARWIERQENDGCWGGIQPPAVYSVIALYLLGYDLEHPVMRAGLE  
SLDRFAVWREDEGARMIEACQSPVWDTCLATIALADAGVPEDHPQLVKASDWMLGEQIVRPGD  
WSVKRPLPFGGWAFEFHNDNYPDIDDTAEVVLALRRVRRHHPERVEKAI GRGVRRNLMGMQ  
SKNGAWGAFDNDTSAFPNRLPFCDFGEVIDPPSADVTAHVEMLAVEGLAHDPRTRRGIQW  
LLDAQEADGWSWFRGWVNYVYGTGSVIPALTAAGLPTSHPAIRRAVRWLESVQNEDEGGWGE  
DLRSYRVREWSGRGASTASQTGWALMALLAAGERDSKAVERGVAWLAATQREDGSDWDEP  
YFTGTGFPWDFINYNLYRQVPLTALGRYVHGEFFAKKRAADAPAEAAPEVKG

>seq\_ID 113

MTDVIDKAVAATGPADPSQGAATLQAAADHLLGLQDDAGWKGELDTNVMTDAEDLLLRQF  
LGRTEEV TREAGDWIRSQQRADGTWANFPDGPADLSTTIEAYTALRMAGDAKDAEHMRAART  
YILDSGGIEASRVFTRIWLALFGEWQSDLPVMPPELIYLPKWFPNLVYDWACWARQTVVPLTI  
VNALRPVRPLGFDLKELRTRRRAPAQGLFSTLDRALHVIYERKPLRSVRDAALRRSADWIIAR  
QEADGWSGGIQPPWYISLMALNLLGYGDVHPVMRKGIEGLDRFTIRDRGRRLLEACQSPVW  
DVTLAMTALRDAELPENHPALVKAADWVLEGEITNPGDWSVRRPRVAPGGWAFEFDNDGYPD  
VDDTAEVVLALNRVAHPDAPAAIRRGVDWLEGMACKDGGYGAFDADNTRTLALKLPFCDFGA  
VIDPPTADVTHTL EAYAALGLANSRASQRALEWLVKAQERDGSWFRGWANHVYGTGAVVP  
AMVAVGVDPEEMIRRAVRWLEEHQNDGGWGEDLRSYRDKSWIGRVSTASQTAWALLAL  
LAAGEERTAVEQGVRFILRTQRADGTWDEHYTGTGFPDFYLNHYLRLVFPISALGRVVR  
AVGAAGDGDAGHAGHAGTVS

## Enzyme Sequences

&gt;seq\_ID 236

MTATTDGGGAI TGGADPRHDS TAAPAAAAAGP SGGGTGLPEGVREAVDRATAELLARQDPAG  
 WWKGDLDQNTVMDAEDLLLRQFLGIRDEAVTRAAALFIRGEQQDGTWATFHGGPPELSATIE  
 AYVALRLAGDPPDAPHMTRASAWIRAHGGIAAARVFTRIWLALFGWWSWDRLELPELPELVFLP  
 PWVPLNIYDFGCWARQTI VPLTVVSALRPVRSAPFALDELHTDARDPVPAKPLPPLASWDGAF  
 QRMDKALHLYRRVAPRRRLKAAMAAAGRWI VERQENDGCWGGI QPPAVYSVIALHLLGYDLDG  
 HPVVRAGLES LDRFAVWREDGARMVEACQSPVWDTCLAAIALADAGLPDHPALVRAADWM  
 LGEEIRRPDGVAVRRPGLAPGGWAFEFHNDNYPD IDDTAEVVLALRRIRHPQPGVEAAIARG  
 VSWTLGMQSKNGAWGAFDADNTPFPNRLPFCDFGEVIDPPSADVTAHVVEMLAEGRAADP  
 RARRGI AWLLAEQEPDGPWFGRWGTNYVYGTGSSVPALTAAGIAPSHPAVRRVAVRWLESVQ  
 NEDGGWGEDQRSYRDRSWAGKGASTASQTAWALMALLSAGERDGA VARGLAYLVETQRP  
 DGTWDEPYFTGTGFPWDFS INYHLRYQVFPPLTALGRYLHGEFFGPERRNVPPAGES

&gt;seq\_ID 134

MSLTSDDPS PATPATQPT SARPSLSDRRSRSGGS AVAGPVLVTRPVAPVAKSGAVTPTATSG  
 AVTSTATSGPALLPDLADLADPTGPLAGAASATVRAAGGAGTRTQQTQGLGSELAGPQAD  
 QVADRAAAVLGRARDHLGLQSEAGWVKGELETNVTMDAEDLMLRQFLGILPPELAAETGRW  
 IRSKQDDGGWPTFHGGPSDLSTTFEAYVGLRLAGDLPDAPHMLAAASFVRAHGGLAATRVF  
 TRIWMALFGEWPEVPLPELVLLPSWVPLNVYDFGCWARQTVVVALTVGHFRPVRSLGF  
 SIDELRVAAVRPDRAPLVSWTGVFQRLDAGLRRYQRHPVKTLRELALRRATEWVLARQEQADG  
 GWGGIQPPWVYSIMALHLMGY SMDHPVLVAALDGLLETFTVREQVREGDEVVTVRRLAECQSP  
 VWDTALAVVALADAGLDARHPAMRKAGEWLVREEVTVPGDWRVRRPNLEPGGWAFEFANDI  
 YPDVDDTAEVVLAVRRLGSGWDDVDP TFAKQARASVERAVNWSVGMRSANGAWGAFDAD  
 NVRELATKIPFCDFGEVIDPPSADVTAHVMEMLADLGRADHPVTQRAVRWLLDDQEPGGSWF  
 GRWGVNHVYGTGAVVPALISAGVAADHPAIRSAVRWLVAHQHPDGGWGEDLRSYQDDAWV  
 GRGEPASQTAWALLALLAADPMNEAVGRGVRLCDTQLPNGTWDEPYTGTGFPWDFSIN  
 YHLRYLVPPLTALGRYVTLTGRSAA

&gt;seq\_ID 225

MTATTDGSGTGAALP PRVTAASD TDIPVAAGVPDIAAAMRRATDFLLSRQSDQGWWKGDLE  
 ETNV TMDAEDLLLRQFLGIRDEGTTRAAALFIRGEQREDGTWATFHGGPGDLSATIEAYVALRL  
 AGDPPDAPHLARASAWIREQGGIAASRVFTRIWLALFGWKKWEDLELPELPELIWPPAVVPLNI  
 YDFGCWARQTI VPLTIVSAERFVRPAPFDELHTDPA RPNPPRALAPVTGWGAFQRLDKAL  
 HVLRGAVPRRLRRAAMNTAARWI IERQENDGCWGGI QPPAVYSI IALHLLGYDLNHPVMRAGL  
 ESLDRFAVWREDGARMIEACQSPVWDTCLATIALADAGLPADHPQLVKAADWMLGEQIVRPG  
 DWSVRRPHLPPGGWAFEFHNDNYPD IDDTAEVVLALRRVAHHDPERVDNAI GRGVRRNLGM  
 QSRNGAWGAFVDNTSPFPNRLPFCDFGEVIDPPSADVTAHVVEMLAEGLAHDPRTRRGVQ  
 WLLAEQEPNGSWFGRWGVNYLYGTGSSVPALTAAGISGSHPAIRRAVAVLESVQNDGGWG  
 EDLRSYRDARGWSGRGASTASQTAWALMALLAAGERESRAVERGVWLAATQHEDGSWDE  
 PYFTGTGFPWDFS INYHLRYQVFPPLTALGRYVNGEPLAGKPRAAGAA TAREDTGQEQLAEAK  
 GS

&gt;seq\_ID 223

MTATTDGSGTGAANI TGAPADDPTDTRTAANDVTDIARRAERSVEHLLGRQDEQGWKGDLE  
 TNVTMDAEDLLLRQFLGIRDEGTTRAAALFIRGEQREDGTWATFHGGPGDLSATIEAYVALRLA  
 GDRPDEPHMARASAWIRQGGIAAARVFTRIWLALFGWKKWEDLELPELPELMFPFKVPLNI  
 YDFGCWARQTI VPLTIVSAKRFPVRPAPFALDELHTDPDHPNPPRKLAPPTSWDGLFQRLDKGL  
 HLYHKVAPRPRRRIAMNVAARWI IERQENDGCWGGI QPPAVYSV IALHLLGYDLNHPVMKAGLA  
 SLDRFAVHREDGARMIEACQSPVWDTCLATIALADAGLRPDHPALVKAADWMLAEEITRPGDW  
 SVRKPELAPGGWAFEFHNDNYPD IDDTAEVVLALRRVRHPDARLEAAIARGVRRNLGMQSR  
 NGAWGAFDADNTSPFPNRLPFCDFGEVIDPPSADVTHGVVEMLAVEGLANHPRTREGIEWLLA  
 EQEACGAWFGRWGVNYLYGTGSSVPALITAGLPAGHPAIRRAVDWLESVQNDGGWGEDLRSY  
 QEKEKWIHGESTASQTAWALLALLAAGRDTASVTRGVTWLTAEQQADGSWDEPYFTGT  
 GFPWDFS INYHLRYQVFPPLTALGRYVHGDPPADR TDAEAGV

&gt;seq\_ID 226

MTATTDGSGTGAALP PRVTAASENDTDIPEAAGVPDIAAHAMRRATDFLLSRQDDQGWWKGDLE  
 ETNV TMDAEDLLLRQFLGIRDEDTTRAAALFIRGEQREDGTWATFHGGPGELSTTIEAYVALRLA  
 AGDPPPEAPHMARASAWIRERGGIAAARVFTRIWLALFGWKKWEDLELPELPELIWPPSWVPLNI  
 YDFGCWARQTI VPLTIVSAKRFPVRPAPFDELHTDPRRPRPPRPHAPPNTWDGAFQRLDRAL  
 HALRRAVPRRVRQAAMNAAARWI IERQENDGCWGGI QPPAVYSV IALHLLGYDLRHPVMRAGL  
 ESLDRFAVWREDGARMIEACQSPVWDTCLAAIALADAGLPADHPSLVKAADWMLGEQIVRPG  
 DWSVRRPHLPPGGWAFEFHNDNYPD IDDTAEVVLALRRVRHHDPERMDSA IGRGVRRNLGM  
 QSKNGAWGAFVDNTSPFPNRLPFCDFGEVIDPPSADVTAHVVEMLAVEGLAHDPRTRRGVQ  
 WLLAEQEPDGSWFGRWGVNYLYGTGSSVPALAAAAGIPGSHPAIRRAVAVLEKVVQNDGGWG  
 EDLRSYRHVREWSGRGASTASQTAWALMALLAAGERDSGAVERGVWLAATQREDGSWDE  
 PYFTGTGFPWDFS INYHLRYQVFPPLTALGRYVHGEPPSKKQTAARNGSAQPLAGVKGSR

&gt;seq\_ID 219

MDPALSRVAVDMLLEHQDPAGWGCGEFETNVTI TAEHILLRFLGLDPSPLRDAVTRYLLGQQR  
 EDGSWALY YEGPADLSTSI EAYAALKVLGLDPTS EPMRRALQV IHDLGGVAQARVFTRIWLAMF  
 GQYPWDGVP SMPPELIWLPSPAPPNLYDFACWARATI TPLLI I LARRPVRPLGCDLGEVLVLP  
 EHLTRVPGSGFPFWGDV LKRYDHLVRHPGRDRACQRI VEWI IARQEADGSWGGI QSAWV  
 MSLIALHLEGLPLDHPVMRAGLAGFDRVALEDERGWRLQASTSPVWDTAWAVLALRRRAGLPR  
 EHPRLALAVDWLLQEQIPGGGDWQVTRGTIPGGGWAFEFHNDNYPD IDDTAVVVLALLEAGH  
 EDRVRNAVERAARWILAMRSTDGGWGA PDRDNAREV IHRLP IADFGTLIDPPSEDEVTAHVLEM

-continued

## Enzyme Sequences

LARLSFPSTDPVVARGLEFLQQTQRPDGAWFGRWGVNYIYGTWCVAVSALTAFTADTADATARAM  
VPRVAWLLDRQNDGGWGETCGSYEDPNLAGVGRSTPSQTAWAVLALQAAGLQGHAPACR  
RGLDFLRERQVGTWEEREHTGTGPPGDFFINYHLYRHVFPPTMALAGAATGMDSPR

&gt;seq\_ID 220

FLGIRDEATTRSAAALFIRGEQREDGTWATFHGGPPDLSTTVEAYVALRRLAGDSPDAPHMTRAA  
HWVRSQGGIAEARVFTRIWLALFGWVWDRLELPELPELIFLPPWAPLNIYDFGCWARQTIIVPL  
TVVSAKRVPVRPAPFPLDELHDTDPADPAPRARFAPLASWNGAFQRLDRALHAYRKVAPRALRRA  
AMATAGRIVERQENDGCWGGIQPPAVYSMIALHLLGYDLGHPVMRAGLES LDRFTLTREDG  
SRMV EACQSPVWDTCLATIALADAGVPADHPQLVRAADWMLDEQIERPGDWSVRRPHLAPG  
GWAFEFHNDNYPDI DDTAEVVLALRRVRHPDTRMERAISLGVRNLMGMQSKNGAWGAFDV  
DNTSSLNRLPFCDFGEVVDPPSADVTAHVVEMLAAEGLAADPRTRRAVDWLLAEQEPGAW  
FGRWGVNYLYGTGSVVPALVADGLPTTHPAIRRAVAVWLESVQNDGGWGEDLRSYREQGRM  
ARGASTASQTAWALLMALLAAGERESRAARRGVTLAETQHEDEGSWEEPYTGTGFPWDFINS  
YHLYRQVPLTALGRYTRGAPEGA

&gt;seq\_ID 125

MQTQNRVTSTQKVELSNLTQAI IASQNYILSRQYPEGYWWGELESNITLTAETVLLHKIWKTDKT  
RPFHKVETYLRRQNEQGGWELFYDGGELSTSV EAYMALRLLGVTPEDPALIRAKDFILSKG  
GISKTRIFTKFHLALIGCYDWKGIPSI PPWIMLPDNFPFTIYEMSSWARESTVPLLI VFDKPIFEI  
EPAPNLDELAYAEVENVKYALPRNHNWSDIFLGLDKLFWTEKNNLVPFHKKS LQAEEKWMLN  
HQQESGDWGGIMPPMVNSL IAFKVLNVDVADPSVQRGF EADRFSEIEEDTYRVQACVSPVWD  
TAWVIRALVDSGLKPDHPSLVKAGEWLLDKQILEYGDWAIKKNQKPGGWAFFENRFPYDLD  
DSAVVVMALNGIKLPDENRKKAAINRCLWEMATMCKPKGGWAAFDVNDQAWINEIPYDGLK  
AMIDPNTADVTARVLEMVSGCLKMDENRVQKALFYLEKQESDGSWFGRWGVNYIYGTSGV  
LSALAVIAPNTHKPKMEKAVNWLISQNEDEGGWGETCWSYNDSSLKGTGISTASQTAWAI IGL  
LDAGEALETLATDAIKRGIDYLLATQTPDGTWEEAEFTGTGFPCHFYIRYHLYRHVPLIALGRY  
WKIGLKTSPVIPLN

&gt;seq\_ID 228

MLARRATDRAVRHLLSRQDEQGWKGDLETNVTMDAEDLMLRHFLGIQNPVLDAAAGRYIRS  
QQAADGTWATFHGGPPPELSATVEAYVALRRLAGDPPDAPHMAAASAWVRNNGGVASRVFTRI  
WLALFGWWRWEDLELPELPEI IYFPWPVLPNLVDFGCWARQTIIVPLTVVSAKRVPVRPAPFSLDE  
LHADPRRPNPRAAPLASWDGAFQRLDRALHLYRKVALRPLRRAALRS CARWIVERQENDG  
CWGGIQPPAVYSVIALHLLGYDLHHPVMRAGLES LDRFAVWREDGSRMI EACQSPVWDTCLA  
VIALADAGLADHPALVKSADWMLAEIEDRPGDWSV KRPR LAPGGWAFEDNDNYPDI DDTAE  
VILALRRVDHPRPERIAAAVRRGVRWTLGMQSRNGAWGAFVDNTSP LPNRLPFCDFGEV IDP  
PSADVTAHVVEMLAHEGGARDPRTRRAVGVWLLAEQEPGAWFGRWGTNYVYGTGSVVPALV  
AAGLPAHPAIRRAVWLESVQNEDEGGWGEDQRSYDPDEWIGHGASTASQTAWALLALLAAG  
ERESKAVERGVWLAATQDQDGSWDEPYFTGTGFPWDFINSYHLYRVLVPLTALGRYVSGEA  
TGARPRRT

&gt;seq\_ID 241

MTATTDGSGTALPPRADAASEHDIETPEAAGVREAAVRAARRATD FLLSRQDAQGWKGDLE  
TNVTMDAEDLMLRQLFGLVDEKTAQAAALFIRGEQREDGTWASFYGGPDELSTTIEAYVALRL  
AGDAPDSPHLAKASAWIREQGGIAAARVFTRIWLALFGWVKWEDLELPELPELIFWPKVPLNI  
YDFGCWARQTIIVPLTIVSAKRVPVRPAPFPLDELHDTDPARPNNPRLPAPAFSWDGAFQRMKGL  
HALRKVAPRGLRRAAMNAARWI IERQENDGCWGGIQPPAVYSI IALHLLGYDLQHPVMREGL  
ASLDRFAVWRBDGARMV EACQSPVWDTCLAAIALVADAGLPADHPQLVKAADWMLGEEIVRPG  
DWSVRRPGLPFGGWAFEFHNDNYPDI DDTAEVILALRRITHHDVPRVDKAVGRGVRWTLGMQ  
SKNGAWAAFDVNTSPPFNRLPFCDFGEVIDPPSADVTAHVIEMLAVEGLAHDPRTRRGI EWL  
LAEQEPDGSWFGRWGVNYVYGTGSVVPALV AAGLPGAHPAIRRAVSWLESVQNDGGWGE  
DLRSYKYVKEWSGRGASTASQTAWALLMALLAAGERD SKAVERGVWLAATQREDGSWDEPY  
FTGTGFPWDFINSYHLYRQVPLTALGRYVHGEPPADRLKGS

&gt;seq\_ID 238

MHEGEAMTATTDGSGTAATPPATTASAPLHLSPEARETHEATARATRAVD FLLARQSDGEGW  
WKGDLETNVTMDAEDLRLRQLGIRDEATTRAAALFIRGEQEDGTWNTFYGGPGLSATIEG  
YVALRRLAGDSP EAPHMRKASAFVRAQGGVARARVFTRIWLALFGWVKWEDLEPEMPPPELMPF  
PKWAPLNIYDFGCWARQTIIVPLTVVCAQRVPVRPAPFALEELHDTDPADPPAPPPVSWDNV  
FHKLDKLLHGYRRI APRRVREAAAMRAATWIVERQENDGCWGGIQPPAVYSI IALNLLGYDL  
HPVLRAGLASLDRFAVWRREDGARMIEACQSPVWDTCLATVALADAGVPADHPQMIKAADWML  
AEQIVRPGDWVVRPDLPPGGWAFEFHNDNYPDI DDTAEVVLALRRVAHPDATRVDKAVRRA  
VDWNVGMQSKNGAWGAFADNTSPPFNRLPFSDFGEVIDPPSADVTAHVVEMLAEGLAHH  
PRTRRGI EWLLKNQEGNSWFGRWGVNYVYGTGAVVPALV AAGLPAHPAIRRSVSWLQGV  
QNEDEGGWGEDLRSYQDSAWHGRGHSTASQTAWALLALLAAGERETEQRVRRGIAYLVETQTE  
DGTWDEPWFTGTGFPWDFIN YHLYRQVFPVPTALGRYLNGTGPGEN

&gt;seq\_ID 237

MRRRSRPGGAGPEADYGPASAPDRLRGDAARGDAARRVQDATARAIRNLLGRQDPAG  
WKGDLETNVTMDAEDLRLRQLGIRDEAVTQAAALFIRREQREDGTWATFHGGPPPELSATIE  
AYVALRRLAGDAPDAPHMMATASAWIRAHGGLAARVFTRIWLALFGWWDWENLELPELVLVLP  
PWPVPLNIYDFGCWARQTIIVPLTVVSAMRPVRPAPFALDELHDTARVVPVPRRMAPPTWNGA  
FQWMDRALHVYRRFAPRRLREAAASAGRWI IERQENDGCWGGIQPPAVYSVIALHLLGYDL  
GHPVMRAGLES LDRFAVWREDGSRMIEACQSPVWDTCLAAIALADAGVPRDHPALVKAADW  
MLGEEIVRTGDWAVRRPGLAPGGWAFEFHNDNYPDI DDTAEVVLALRRIRHPDPAVREAAIAR  
GVSWNLMQSRGGAWGAFADNTSPPFNRLPFCDFGEVIDPPSADVTAHVVEMLAEGRRA

-continued

## Enzyme Sequences

DPRTRRGI A W L L A E Q E P E G P W F G R W G T N Y V Y G T G S V V P A L T A A G L S P G H P A I R R A V L W L E S V  
Q N P D G G W G E D Q R S Y Q D R A W A G K G E S T P S Q T A W A L M A L L S A G E R D A K T V E R G I A Y L V E T Q L A  
D G G W D E P H F T G T G P P W D F S I N Y H L Y R H V F P L T A L G R Y L Y G E P F G H D G R H I G A H L G D R T G V P A  
E G V

&gt;seq\_ID 239

M D F L L D R Q S D E G W W K G D L A T N V T M D A E D L L L R Q F L G I R D E A T T Q A A A L F I R G E Q Q E D G T W N T  
F Y G G P D L S A T I E G Y V A L R L A G D S P E A P H M R K A S A F V R A R G G V A R A R V F T R I W L A L F G W W K W  
E D L P E M P P E L M F F P K W A P L N I Y D F G C W A R Q T I V P L T V V C A Q R P V R P A P F A L E E L H T D P A D P N P  
A Q P A P P V A S W D N V F H K L D K M L H G Y R K V A P R R V R E A A M R A A A T W I V E R Q E N D G C W G G I Q P P A  
V Y S I I A L H L L G Y D L D H P V R A G L E S L D R F A V W R E D G A R M I E A C Q S P V W D T C L A T V A L A D A G V P A  
D H P Q M I R A A D W M L A E Q I V R P G D W V V R R P D L P P G G W A F E F H N D N Y P D I D D T A E V V L A L R R V A H  
P D A T R V D K A V R R A V D W N A G M Q S K N G A W G A F D A D N T S P F P N R L P F C D F G E V I D P P S A D V T A H  
V V E M L A E E G L A H H P R T R R G I E W L L E N Q E A N G S W F G R W G V N Y V Y G T G A V V P A L V A A G I P A A H P  
A I R R S V S W L G Q V Q N E D G G W G E D L R S Y Q D T A W H G R G H S T A S Q T A W A L L A L L A A G E R D S E Q V  
R R G I A Y L V E T Q T E D G T W D E P W F T G T G P P W D F T I N Y H L Y R Q V P P V T A L G R

&gt;seq\_ID 235

M T Q T V P R T A A S A P A A R T A A D T V A A A V Q F L R R E Q D R A G W W K G E L A T N V T M D A E D L L L R H F L G I  
L T P Q I A E E S A R W I R S Q Q R A D G T W A N F P D G P A D L S T T V E A W V A L R L A G D P A D A P W L A T A A E W I  
R E H G G I E A T R V F T R I W L A M V G Q W S W D D L P S L P P E L I F L P S W F P L N V Y D F A C W A R Q T I V P L T T I V G  
T L R P A R K L P F P V A E L R T G K R P K P R A P W T W D G V F Q N L D T A L H A Y A K L P L N P V R K L A L K Q A A E  
W I L A R Q E A D G S W G G I Q P P W V Y S I L A L H L L G Y S L D H P A L K A G I A G L D G F T I R E K T D Q G W R R L E A  
C Q S P V W D T A L A M T A L L D A G V S P G D E S L V R A A E W M L G E E I R V P G D W A V R R P S L K P G G F A F E F A  
N D G Y P D T D D T A E V V L A L R R M G K P D H L R I R E A V D R S V A W L E G M Q S S D G G W G A F D A D N T Q V L T  
T R L P F C D F G A V I D P P S A D V T A H V V E M L A A E G K A D T R E C R R G I R W L W D N Q E A D G S W F G R W G A  
N Y V Y G T G A V V P A L V A A G V P G T D P R I R R A V R W L A E H Q N D D G G W G E D L R S Y D D R S W A G R G D S  
T P S Q T A W A L L A L L A A G E R E S T V V A R G V E W L C E R Q R P D G G W E D E D K H T G T G F P P G D F Y L S Y H L Y  
R V V F P L S A L G R Y V R G G S

&gt;seq\_ID 159

M S G Q S N F T G G K K M T P A E G S S P A P A L L E K A A P S I E L D E R S D P L S R T L A R A V S W L V A A Q D G A G  
H W V A P L E A D A T I P S E Y V F L H E V L G R P L D P V R R D K I V R A I L S V Q G K E G A W P L F H D G D P D I S A T V K  
A Y Q A L K L C G F D P S H P A L V R A R E W L S Q G G A G K V N V F T R I A L A I F G Q Y S W T K I P A P A E M V L L P S  
W F P F S I Y S V S Y W S R T V I V P L L F I Y H H K P L V R L S P E R G I S E L F D P A R P D G E S F A P S P D F F S L R N L F L  
L L D K V L Q V W N R H P P G F L R K K A L S F A M E W M V P R L K G E G L G A I Y P A M A N S A V A L S L E G Y E L D H  
P L M Q R V L A S I D D L L E G E K E V L V Q P C V S P V W D T A L A M G A L I E A G I S P D S P T V D R A M E W F C A R E V  
R T R G D W A I R A P D C E P G W A F Q F E N D Y Y P D V D T A M V L M G M A K I L P A R P D L A A R M E G V F R R A  
T L W V M A M Q G T D G G W G A F D R D N D L L F L N H I P F A D H G A L L D P S T A D L T G R V L E L L G A L G Y G P D F  
P P A A R A I R Y L R R E Q E E D G S W F G R W G V N Y I Y G T W S V V A G L K S I G V P M S E P W M R S M E F L L A R  
Q N P D G G W G E D C L S Y A S R D F A G R G A S T P S Q T A W A L I A L L H G H A G H M A V R Q G V D Y L I Q Q M T P  
E G T W N E E L F T G T G P P R V F Y L R Y H M Y R H Y F P L W A L A L Y R N M T E R G A L G H E R V D F W K T A P Y A  
P I A R S V

&gt;seq\_ID 232

M T A T D G S T G A L P P R A P S A S D T H G T P V A A G V Q E A A L H A V G R A T D F L L S R Q D A Q G W W K G D L  
E T N V T M D A E D L L L R Q F L G I R D D A T T R A A A L F I R G E Q R P D G T W A T F Y G G P D L S A T V E A Y V A L R L  
A G D D P A A P H M A K A S A W I R A R G G I A A A R V F T R I W L A L F G W W K W D D L P E M P P E I V Y F P T W M P L N I  
Y D F G C W A R Q T I V P L T V V S A K R P V R P A P F P L D E L H T D P G R P N P P R L D R L G S W E G A F Q R L D R A  
L H G Y H K V A L K R L R R A A M N R A A R W I V E R Q E N D G C W G G I Q P P A V Y S V I A L H L L G Y D L D H P V M R A  
G L E S L D R F A V W R E D G A R M I E A C Q S P V W D T C L A T I A L A D A G L P P D H P Q L V K A A D W M L G E E I V R P  
G D W S V K R P Q L P P G G W A F E F H N D N Y P D I D D T A E V V L A L R R V R H P D P E R V E R A V R R G V R W T L G  
M Q S G M G A W A A F D A D N T S P F P N R L P F C D F G E V I D P P S A D V T A H V V E M L A A E G L S H D P R T R R G I  
E W L L A E Q E P G G A W F G R W G V N Y V Y G T G S V V P A L V T A G L P A A H P A I R R A V A W L E T V Q N D D G G  
W G E D L R S Y P D P A E W G G K G A S T A S Q T A W A L L A L A A G E R D G K A T E R G V A W L A R T Q R E D G S W  
D E P Y F T G T G P P W D F S I N Y H L Y R Q V F P L T A L G R Y V H G E P A V L K P G T R

&gt;seq\_ID 224

M T A T D G S T G A N L R A A A A S D P T E S T S A A P D M M A V A R H A A E R S V E H L L G R Q D E Q G W W K G D L  
A T N V T M D A E D L L L R Q F L G I Q D P E T V K A A A R F I R G E Q L G D G T W N T F Y E G P P D L S A T V E A Y V A L R L  
A G D R P D D P H M I R A A G W V R E Q G G I A E S R V F T R I W L A L F G W W K W D D L P E L P P E L M F F P K W V P L  
N I Y D F G C W A R Q T I V P L T I V S A K R P V R P A P F A L D E L H T D P A C P N P S R P T A P A A S W D G V F Q R L D K A  
L H L Y H K V A P R R L R R I A M N E A A R W I I E R Q E N D G C W G G I Q P P A V Y S V I A L H L L G Y D L D H P V M R A G L  
E S L D R F A V W R E D G A R M I E A C Q S P V W D T C L A T I A L A D A G V S P D H P A L V R A A D W M L G E E I V R P G  
D W A V R K P G L A P G G W A F E F H N V N Y P D I D D T A E V A L A L R R V R H P D P A R V D A A I E R G V R W N L G M  
Q S R N G A W G A F D A D N T S P F P N R L P F C D F G E V I D P P S A D V T G H V V E M L A V E G R A H D P R T R R G I  
E W L L A E Q E A S G A W F G R W G V N Y I Y G T G S V V P A L I A A G L P A A H P S V R R A V D W L R S V Q N D D G G W  
G E D L R S Y R E E K W I H G S S T A S Q T G W A L L A L L A A G E R E T R S V E R G V A W L A T Q Q A D G S W D E P  
H F T G T G P P W D F S I N Y H L Y R Q V F P L T A L G R Y V Y G D P F A T A T A I G A G T G K G A

&gt;seq\_ID 243

M S I S A L Q T D R L S Q T L T Q S V V A A Q Q H L L S I Q N P E G Y W W A N L E S N A S I T A E V V L L H K I W G T L D S Q P  
L A K L E N Y L R A Q Q K T H G G W E L Y W N D G G E L S T S V E A Y M G L R L L G V P A S D P A L V K A Q Q F I L H R G G  
V S K T R I F T K F H L A L I G C Y R W Q G L P S L P A W V M Q L E S P F P F S I Y E L S W A R G S T V P L L I V F D K K P V Y  
P L Q P S P T L D E L F T E A S E A N V R W E L E K G D W S D A F L W L D K A F K L A E S V D L V P P F R E S I R K A E K W V  
L E R Q E P S G D W G G I I P A M L N S M L A L R A L G Y S V S D P V R R G F Q A I D N F M V E S E T C W A Q P C I S P V  
W D T G L A V R S L T D S G L S P N H P A L V K A G E W L L D K Q I L S Y G D W S V K N P Q G Q P G G W A F E F E N S F Y

-continued

## Enzyme Sequences

PDVDDTAVVAMALQDITLPENEPLKRRAIARAVRWIATMQCKTGGWAAFDINNDQDNLNDI PYG  
DLRAMIDPSTADITGRVLEMHGRFAADLDLANSYAADLSPYRLSRGLNYLI KEQELDGSWFGR  
WGVNYI YGTGQALSALALIAPERCR I Q I ERGI AWVSVQNADGGWGETCESYKDKSLKKGKI ST  
ASQTAWALLGLLDVFCFLDPAKIAVDRGIQYLVSTQSEGTWQEESTGTGFPQHFLYRLRYLY  
CHYFPLMALGRYQVINSSAGI

&gt;seq\_ID 197

MTSGTFGAKRVDLLAAFEHSAPAETRETCTCVGLQTAIARTRQYLLDQOHSEGFVVALEEGDTIL  
ESEYILLLAFLENEGQSPDAQAAARYLLTKQNTDGSWSNFPGGPIDVSCAVKAYLALRI TGHAA  
EPALIRAREAILQAGGVERVNSFTRFYLAMGLIPIYSLCPAVPPEVLLPDWFPINLSQMSAWSR  
TIVVPLSLLWAFQPAVELNDADGHQITIEELYASPEKQLPRFIRGVNHESNSNGMWNWSRFFFR  
VDQCLKSI ESYGIKPLRSRAVRKCVQWILDRQEMSDGLGAI FPPPIVWTL IGLKCAFPDQHPMV  
KQQRDELNRLMLREQDALRLQPCLSPVWDTAISIALRESGVEPDHPALSKARNWLLSKEVRHA  
GDWSKAHPETPVSGWYFENNEFYPDVDDTAMVLI ALASTLPEEATPLAISHGVLPVQTGWSA  
ESTSRVQALKQLENHRPVL EAMGRGVQWLKALQSKDGGWGAFFSD INKELLTKVPFADHNAM  
LDETNA DI SARVLEAYAAVGI SFNDPSVQRALEF IWNDQEDDHAWYGRWGVNYI YGTWQV LV  
GLTAIGISAHDPRLVRAAGWLKSKQACGGWGETPATYDNPTLRGQGTPTASQTAWAVLGLTIA  
AGEQNSIECQRGVEFLKTKQKHNGTWDEEEFTGTGFPVRYLYHYLYPLFYPLMALGRFARA  
GGRVNFAG

&gt;seq\_ID 158

MTTNAAT SARSGEDAIRVSGQOLETAIASARNL LALQRPDGHFVFELEADATI PAEYVLMR  
HYLAEPVDAVLEEKIARYLRR IQSDDGGWPLFRDGASNISASVKAYYALKMIGDAPNAPHMQKA  
RAWILAQQGASHSNVFRNLLALFGAIPWSGVVMPVEIMLLPKWFPFHIDKISYWARTVLIPLT  
VLNALKPVARNPKGVGIAELFVTPPDQVRNWKGPQKFPWSQVFGGIDRVLRLPEPAPFKSL  
RKKSIDKAVAFATERLNGEDLGGI F PAMVNALLVYDALGYPHDHPDVTARGSEIEKLLVIKDD  
AYCQPCLSPVWDTALAVHALMESGVAQADQNVDRALAWLKPQLVLDTVGDWAASRPGRVPG  
GWAFQYANAYYPDVTAVVMMAMDRAGGDAAKRDHYRESMARGREWVAVGQSKNGGW  
GAFDADNTYEYLNQIPFSDHGALLDPP TADVSARCVSMLAQLGERRETSVPLDKAMRYLESTQ  
EKDGSWYGRWGMNYI YGTWSVLCALNAAGVAPSA PSMRKAADWLLSI QNSDGGWGEDGES  
YSLDYKGYEPAPSTASQTAWALLMGLMAAGEVDHPAVQRGVAYLAAKQSGDGFWEERFTAT  
GPRVRYLYRHYGSKFFPLWALARYRNLSNAANSKSVLVGM

&gt;seq\_ID 77

MAADGSALSESRLSSEALDRVLSAHTALSQAQQDDGHVVELEADATI PAEYILLEHFMDRID  
DALEQKIAIYLRRIQSEEHGGWPLYHNGKFDLSATVKAYFALKAVGDDINAPHMQRAREAILDH  
GGAERSNVFTRSQLALFGEVWRATPVMPVLEMLLPKAFPSVNMMSYWSRTVIAPLLVLAAL  
RPVAANPRQVHVRELFVTPPEKVDWI RGPYRSANGYVFKGLDSVLRPVVPIPEKTHKKAIQ  
AALDFI EPRLNGKDLGAI YPAMANVMMYRAMGVPEDEDPRAKTAWEAVQALIVEKDD EAYC  
QPCVSP IWD TGLSGHAMIEAASGPNGI APEKTVAELKKASAWLRSKQILNVKGDWAVRNPILA  
PGGWAFQYGN DYYPDVTAVVGMMLLHREGDPTNAEAI ERARTWIVGMQSTDDGGWGAFFD  
NNKDVLNHI PFADHGALLDPP TADVTARCI SFLAQLRNPED EPIV IQRGLEYLRKEQEKDGSWGF  
RWGTNYI YGTWSALCALNAAGVSHDDPAVV KAVEWLRVSVQRADGGWGECSYEGGPHGT  
YGESLPSQTAWAVLGLMAAGRDDPAVTRGIAWLADQQDANGEWHEDPYNAVGFPPKVFYLR  
YHGYKQFPPLMALARYRNLESNTRRVSFSG

&gt;seq\_ID 6

MTVSTSSAFHHSLSLDDVEPI IQKATRALLEKQHQDGHVVELEADATI PAEYILLKHYLGEPE  
LEIEAKIGRYLRRIQGHEGGWSLFGYGGDLDSATVKAYFALKMIGDS DPADPHMLRARN EILARG  
GAMRANVTRIQQLALFGAMSEHVPQMPVLEMLMPEWFPVHINKMAYWARTVLPVLLVQLAL  
KPVARNRRGILVDELFPDVLPTLQESGDP IWRFFSALDKVLHKVPEPYWPKNMRKAIHSCV  
HFVT ERLNGEDLGAIPAIANSVMYDALGYPENHPERAIARRAVEKLMVLDGTEDQGDKEV  
YCPCLSP IWD TGLVAHAGMLVEVGGDEAEKSAI SALSWLKQQLI LDVKGDWARRPDLRPGW  
AFQYRNDYYPDVTAVVMMAMDRAAKLSLHDDFEESKARAMEWTIGMQSDNGGWGAFDA  
NNSYTYLNNI PFADHGALLDPP TADV SARCVSMMQAAGISITDPKMKAAVDYLLKEQEEEDGSW  
FGRWGVNYI YGTWSALCALNAALPHDHLAIQKAVAWLKNIQNEDGGWGENCDSYALDYSGY  
EPMSTASQTAWALLGLMAVGEANS EAVTKGINWLAQNQDEEGLWKEDYSGGGFPVRYLY  
RYHGYKQFPPLWALARYRNLSNAANSKSVLVGM

&gt;seq\_ID 89

MNDL TNSAPGARPDATP SAAGPTPAEAGGAVAPSRAVQPADTQTAATGAAGAAA AVGAT  
PAELAATAPASSGT PAGASAAPAPSGT PVDPAELASAAAPAPSGAT PAATATAATAPAPARA  
ASIDAPALAAADLDAI TRATDALLAAQQADGHWIYELEADSTIPAEYVLLVHYLGETPNLELERK  
IARYLRRVQLPGGGWPLFDGAPDVSA SVKAYFALKMIGDANAHEMVRARNAIHAMGGAM  
SNVFTRIQLALFVVPVAVFVMMMPVEIMLLPQWFPFHLKSVSWARTVTVPLLVLSAKRPLARN  
PRGVRVDELFPVAPVNAAGLPRAGHS PAWFACFRLDLGLRLTDGLFPRYTRERAIRQALQF  
VDERLNGEDLGAIPAMANSVMYALGYPEDHPNRATARRAIEKLLV IHDD EAYCQPCLS  
VWDTSLAAHALLETGEPRAEAAAIRGLDWRPLQILDVVRGDI SRRPDPVPPGGWAFQYANPH  
YPDVTAVVTLAMDRVAKLAQTDA YRDAIARAREWVGMQSDGGWGAFFEPENTHQLNSI  
PFSDHGALLDPP TADVSGRCLSM LAQLGETAANSAPARRALDYLLAEQDAGGSWYGRWGMN  
YIYGTWSALGALNAAGLFPDPRVKRAAQWLLSI QNPDGGWGEDGDSYKLDYRGYERAASTA  
SQTAWALLGLMAAGEVEHPAVARGIAWLAQQREHGLWDEARFTATGFPVRYLYRHYGKRF  
FPLWALARYRNLRRTGTRRVTVGM

&gt;seq\_ID 201

MLPYNQNSYKEALHGGHAAHNPPTLEEAIKRSQEFLLAHQHPGFWGDLECNVTSASHTLIL  
YKILGIADRYPLHKFEKYLRRMQCSHGGMWMSFGDGYLSATIEAYICLRLNVPQSDPALQRA

-continued

## Enzyme Sequences

LKNILARGGVTKARVFTKVCALLGGFDWAALPSLPPWMLFPWFPPWNIYEASWARGCVVP  
LIVLLEKKPVFQVKPEVSEFDELYVEGRAHACKALPFSAHDWVSNIFVAADRAFKLIMERFGAVPF  
RQWSIKEAKKWVLDROEMGDFI GYNPMLYFAVCLKLWGYEVDPLLQRRALLAHKCLTVETE  
DECWLQSSQSPVWDALVI PALVESGLPPDHPALQKAGQWLEKQILKKGDWALKTGGGRMQ  
DDIGGGWAFQFVNSWYPDVDDSAAVVIALNCIKMPDEDVKNGAIARCLKWI AFMQGRNGWA  
AFDRDSNRQWMDATPFSDI EAMLDVSTADVTRVLEMVGLMRLKHAAPANNLSGKAHRHIS  
TESIARGVDYLTKEQEKEGCWGRWGVNYI YGTRGALMGLSQVAAKTHKKEIARGAAWLVK  
QNKNEKKQGAQDGGWGEACFSYDDPATKQNSRSTASQTGWAMQGLLAAGEVLGRKYEM  
EAVEEGVQFLLDTQRKDGWSSEAEFTGGGFPKHYYLKYHYFAQHFPPLSALARYRARLLQLSR  
PKNQA

&gt;seq\_ID 183

MDGSQRISDMSQQPEGIAVSDEISSAYSVSSLNQDEINVDELENKLTQARSAMLSLQKPDGHW  
CFPLEADCTIPAEYILMMHFMDIDVILENKIARFIREKQDLTHGGWPLYGGAFDISCTIKSYIA  
LKLVDSPDAAHMVRAREAILERGGAAKANVTRLLLAMYEQIPWSGVPPVTELMLLPSWFP  
PHISKVSWSTRVMIPLSLCTIKARA INPRNVDIRELFI VPPBEQEKNYFPQADTWLKRAPMLVER  
VLSRVEPKLPQAIRGYSIRKAEENWTLERLNGECGIGAI FPAMVNAHESLALGAYDHPSRVQC  
RNALRGLLVDEGERAWCQPCTSPVWDTVLTCLALQEDPAADQGPVLKALDWLVDQVLDPEP  
GDWRDKRPDLGGGWAFQYANPHYDLDLDDTAAVAWALDQSDAQRYSKPLDRAANWLAGMQ  
SRNGGFAAFDIDNTYHYLNEIPFADHGALIDPPTSDVTARCVGLLKYKQHQREVWRGISTFLLR  
EQEKNGSWFGRWGTNYI YGTWSVLEAPQLANPDMQHTSVRRRAVKWLESVQRVDDGGWGETN  
DSYLDIQLAGQFPQTSTTFQTAWAVLGLMAAGEVNSKSVRRGINYLHNLQADDHLWEDPWFT  
APGFPRVYLYRHYGSKFFPIWALVRYRALTKERVS

&gt;seq\_ID 102

MNDLSQTPDLDAVLEAADAASNLAEAAVVANAPAVADALATATSPMQTAGASPLDVSITRA  
TDAILAAQQPDGHWIYELEADATIPAEYVLLVHYLGETPNLELEQKIARYLRRIQLPNGGWPLFT  
DGALDISASVKAYFALKMIGDPVDAEHMVRARDAI LAHGGAEHANVPTRIILALFGVSWRAVP  
MMPVEIMLLPMFPFHLKSVSWARTVIVPLLVNNAKRLARNPVKRVIDELFRGAPVNTGMN  
ERAPHQHAGWFGFRCDVTVLRAVDGLLPKASRERAIRAAVAFVDERLNGEDGLGAIFPAMAN  
SVMMYDVLGYPADHPNRAI ARKSLDKLLVI KEDEAYCQPCLSPVWDTSLVAHALLETREAREAE  
QAAERGLAWLRPLQILDVRGDWISRRPNVRRPGWAFQYNNAHYPDVEDTAVVAMAMHRSAA  
LTKSDVDREALARAREWVVMQSSSEGGWGAPEPENTQYLLNMI PFSHAAALDDPPTADVSGR  
CLSMFAQIGELPQNSEPAQRAFDMYMLQEQESDGSWYGRWGLNYI YGTWTALCSLNAAGMSH  
DDPRMRAVQWLVS IQNEDGGWGEAGESYKLDYRGERAPSTASQTAWALLGLMAAGEVD  
HDAVARGIDYLQREQREHGLWDETRFTATGFPRVYLYRHYGSKFFPLWALARFRHLKRNGL  
TRVTVGM

&gt;seq\_ID 90

MIRPMKNSDLPPLSLLDAAILRGRDALAQRQASADGWSWCFELES DATI TA EYILMMHFMGKIDEA  
RQARMARYLRGIRLATHGAWDLVYDGA PDVSCSVKAYFALKAAGDSEDPHMARARETI LKL  
GGAAKSNVTRILLATFGQPWRATPFMPVEFVLPFKWVPI SMYKVA YWAR TTMVPLLVLCSL  
KARAKNPRNVSIRELFTVTAPEAERHYFARGGFVRNLFLGIDRALRPLDALI PKALRRRAIRHAEA  
WCAERMNGEDGMGGIFPPIVYSYQMMDLVGYPEDHPLRRDCENALDKLLVERPDGSVYQCP  
CLSPVWDTAWSTMALEQARAVPDRDAPPVSDAQLRQCI AASYEWLAGKQVTVQVRGDWVEN  
APAA TPAGGWAFQYANPHYDIDDSAVVAAMLHRRGRLLARSTGTDPYAQVVARGLDMWRG  
LQSRNGGFGAFDADCDRLYLNLIPFADHGALLDPPTEDVSGRVLCLGVTGRDEDKPALARAIE  
YVKMRQADGCGWGWGTNYI YGTWSVLAGLALAGENPSQPYIARAI AWLRACQNAAGGW  
GETNDSYLDPALAGTNGGESASNVTAWALLAQMAFGDQWQSESVQRGIRYLLSVQQADGFWW  
HRSHNAPGFPRYLYLHYGYTAYFPLWALARYRRLSQAGAARDVTDGAALAAS

&gt;seq\_ID 167

MREAAVSKVETLQRPKTRDVS LDDVERGVQSATRALTEMTQADGHICFELEADATIPSEYILFH  
QFRGTEPRPGLLEAKIGNYLRR TQSKVHGGWALVHDGPFDMASVVKAYFALKMIGDDIEAPHM  
RAVRKAILQRGGAANANVTRILLALYGEVPPVAVPVMPEVMHLPKWFPPHLDKVS YWARCT  
MVPLFVIQAKKPRAKNPRGVVAELFVTPPDSVTRTWPGSPHATWPTPI FGGIDRVLQKTQDH  
FPKVPQRRAIDKAVAWVSERLNGEDGLGAIFPAMVNSVLMYEVLYGYPPEHPQVKIALEAIEKLV  
AEKEDAYVQPCLSPVWDTALNSHAMLEAGGHQAEANARAGLDWLKPLQIILDKGDWAETKP  
NVRPGGWAFQYANPHYDLDLDDTAVVMMAMDRAQRQHGLVSGMPDYSIESIARAREWVEGLQ  
SADGGWAAAFDADNHHYLNHI PFSHDGALLDPPTADVTRVVSMLSQLGETRATSRALDRGV  
TYLLNDQEKDGSWYGRWGMNFI YGTWSVLCALNAAGVDPQSP EIRKAVAWLIRIQNPDGGWG  
EDASSYKLNPEFEPGYSTASQTAWALLALMAAGEVDDPAVARGVNYLVRTQGGDGLWSEER  
YTATGFPRVYLYRHYGYPKFFPLWAMARFRNLKRNRSRQVQFGM

&gt;seq\_ID 133

MTTDTETALAAAGTPKAAFAPAPRGAADDLVARTVAVEAPPSPAPASDDTLARAVHLKSLQDE  
AGWKGDLLETNTMDS EDMLRHLWGIWNPQAERTARFIRSKQYADGSWPIYHAGPGDLN  
ATVESYVALRMVGDSPQDPHMRAAAAWARARGVPATRI FTRIWALFGWWRWEDLPVLP  
ELIFVPAKMPLSIYKFAWSGRQTVVAIMVLAHRPAGTPPPPIAELFPPTATKKAQAQRKAQKK  
AGHAGGPTAWRDSIDDMFTEPAPGDTL RQPAALAIGPARPAPAKGRRGKQPAAPDVMG  
RAKDDGGGGLPLPARLVSRVGFRTRRALRQAALDHVNNLLFGGIDRFLHVYHRHP IRPVRS  
ALGLAERWIVVRQEDGCGG IQPPTVYSIMALRVLGYPMDHPVMTAALRSLDEYSVTLDPGA  
RMQEQACQSPVWDTCLATIALADAGVPRDDPSLVRADWMLAEVRERRRGDWSVPIPDVPTG  
GWSFEFNDNTYPDVDDSAEVMLALMRVAHPREKVVAAATYRGLQVWVGMQCADGGWGAFFD  
VDNAGELVYKIPFADFGMLTDPSSADVTAHVVELLGGELGLDPRTKRGVEWLLHSQEAEDGS

-continued

## Enzyme Sequences

WYGRWGVNHLTYGTGGVVPALRAAGLPASHPAIQRAADWLVAQKQNDGGWGESCSYSDMS  
TAGVGVSTASQTAWALLALIAAGRVGDGVTGEAARGVAVLAETQTAEGTWDDEDYFTGTGFA  
GYFYINYHLRYRLVWPMALGRYQAALAGKGH

&gt;seq\_ID 7

MNPVVHNLTRPHRSAPRPSALQRSIAAAQAALLQHQAADGHWCPEFEADCTIPAEYILMMHY  
MDERDAALEAKMAAYLRKQENHGGWSLYHGGHFDMSASVKAYFALKLAGDDPEAAHMRRA  
RSAI LAHGGAEANRVFTRI TLALFGQVPWRVPP I PVE I L L F P R N F P M H I Y K V A S W S R T V M V P L F  
I L C S L K P Q A K N P L G V H I R E L F T R P P E D I D D Y F A H A L Q G W V S R I F L W F D R L G R A L E S W I P Q A L R R R  
A I A R A E A W F I E R L N G E D G L N G I F P A M V N A H E A L A L L G Y A A E H P Y R Q Q T R A A L T K L V V E R A G E A Y  
C Q P C V S P V W D T C L A L H A L L E A D G D V S E A A R R S M Q W L L D R Q I T D A P G D W R E R R P H L A G G W A  
F Q Y A N P Y Y P D L D D T A A V A W A L A R A R R P E D R P A V E R A A N W L A G M Q S R N G G F G A Y D V D N T Y Y Y  
L N E I P F A D H K A L L D P P T A D V S G R V L A F L A I L D R E Q D A P V R A R L I Q Y L L R E Q E P S G A W F R W G T N  
Y I Y G T W S V L M G M A E L R D P G A E V R D A M A R A A H W L R S V Q Q D D G G W G E S N D S Y A D P G L A G L G Q  
E S T A A Q T A W A C L A L M A A G D S D S E S L R R G I Q W L Q R H Q E Q P G D W Q D P Y F N A P G F P R V F Y L T Y H  
G Y K I Y F P L W A L A R Y R N I T E R H C A

&gt;seq\_ID 190

MALSNGEIREEIQRLSEELIQRQEPDGSWRFCFENGITIDACTIILLRNLNVDKEELIRQLHDIRIVA  
AQPPGECWRWYHDDKEGHLSATVEAYYALLCSGYSRPEDEPIQRAKRYILDRGGIGQARS LF  
TKAILAATGQRKWPASLSLPIEILLPELPLNFYDFSGYSRVHLVPLLI MAERNFRTRSVRTPDL  
SELF LDARNGEEDPLTLTPESREPLKLIQSGLAHLVGT PRRIRQA AVNR AEQYMLDR IEGDGT L  
YTAGCTVLMVMAELRALDGYEPQHPVIQRAVEGLSQMKFTVDSTGQGGTRYVTIQNSPSTVWDT  
ALISYALQEQAGVSSHPAIQRAADYLRNRQHRRPGDWQIHNPGIVPGGWGFSETNTFVDPVDD  
TTAALRALALHGS EPAVLGAWNRGLNWWVMQNNDDGGWPAFEKNTNKEMLTWLAI EGAKS  
AATDPS EADLTGRTLEYLGNFAKLSVRQDQVARGADWLLSHQEQADGSWYGRWGI CYIYGTW  
AALTGLMAVGMPADHPGIAKAAANWLRIRIQNADGGWGESCRSDQVRRYVPLHASTPSQTAWAL  
DALIAVHRRRAPEIERGVARLIALLHEDDWPSTYPTGAGLPGYFYVHYHSYRYIWP L L A L S H Y V  
NKYGDSSP

&gt;seq\_ID 45

MSGVLLYDKVREEIERRTTALQTMQRQDGTWFCFEGALLTDCHMIFLLKLLGRNDEI EPFVKR  
LASLQTNEGTWKLYEDEDGNNLSATIQAYAALLASEKYSKEDINMRAEMPIKEHGGVSR AHF  
MTKFLLAIHGEYEPALPHFPPTLFLQDSDPLSIFGLSSSARIHLIPMMI CMNKRFRVEKLLPNL  
NHIAGGGQWFRREERSPLFQSFVGDVKKVIAYPLSLHHKGYEEVERFIMKERIDENGLTYSYASA  
TFYMIYALLALGHSIQSPIIEKAVI GLKSYIWKMDRGSHLQNSPSTVWDTALLSYSLQEQANVMKE  
NKMIQKATEYLLRQQT KRMDWSVHAPSIMAGGWGFSVDVNTTIPD VDDTTAALRALARSRG  
SRVDSAWGRGVEWVKGLQNNDDGGWGA FERGVTSRILANLPIENASDMI TDPSTPDI TGRVLEF  
FGTYAPNELPEEQKKAVKWLMDVQELNGSWYKWKGI CYIYGTWAMTGLRALGVPSHPSL  
KKAASWLEHLQYEDGGWGES CQSSVEKKFISLPFSTPSQTAWALDALISYDQETPI IRKGISY  
LLAQSTMNEKYPTGTGLPGGFYIRYHSYGHYIPLLA LAHYI KKYK

&gt;seq\_ID 53

MSGVLLYDKVHEEIERRTTALQTMQRQDGTWQFCFEGALLTDCHMIFLLKLLGRNDEI EPFVKR  
LVSLQTNEGTWKLYEDEDGNNLSATIQAYAALLASERYSKEMNMRRAEMPIKEHGGVSR AHF  
MTKFLLAIHGEYEPALPHFPPTLFLQDSDPLSIFGLSSSARIHLIPMMI CMNKRFRVEKLLPNL  
NHIAGGGQWFRREERSPLFQSLGDVKKVI SYPLSLHHKGYEEVERFMKERIDENGLTYSYASA  
ATFYMIYALLALGHSIQSPIIEKAVTGLKSYIWKMDRGSHLQNSPSTVWDTALLSYSLQEQAVTN  
ENKMIQRATEYLLRQQT KRMDWSVHASSLVAGGWGFSVDVNTTIPD IDDTAALRALARSRG  
DRVDDAWGRGVEWVKGLQNNDDGGWGA FERGVTSKLLSNLPIENASDMI TDPSTPDI TGRVLE  
LFGTYAPNELLEEQKKKAIKWLMDVQEQNGSWYKWKGI CYIYGTWATMTGLRALGVPSHPS  
LKAASWLEHLQYEDGGWGES CQSSVEKKFISLPFSTPSQTAWALDALISYDQETPI IRKGIS  
YLLAQSTMNEKYPTGTGLPGGFYIRYHSYGHYIPLLA LAHYVKKYK

&gt;seq\_ID 44

MSGVLLYDKVHEEIERRTTALQTMQRQDGTWQFCFEGALLTDCHMIFLLKLLGRNDEI EPFVKR  
LASLQTNEGTWKLYEDEDGNNLSATIQAYAALLASEKYSKEDMNRRAEMPIKEHGGVSR AHF  
MTKFLLAIHGEYEPALPHFPPTLFLQDSDPLSIFGLSSSARIHLIPMMI CMNKRFRVEKLLPNL  
NHIAGGGQWFRREERSPLFQSLGDVKKVI SYPLSLHHKGYEEVERFMKERIDENGLTYSYASA  
ATFYMIYALLALGHSIQSPIIEKAVTGLKSYIWKMDRGSHLQNSPSTVWDTALLSYSLQEQAVTN  
ENKMIQRATEYLLRQQT KRMDWSVHASSLVAGGWGFSVDVNTTIPD IDDTAALRALARSRG  
DRVDDAWGRGVEWVKGLQNNDDGGWGA FERGVTSKLLSNLPIENASDMI TDPSTPDI TGRVLE  
LFGTYAPNELLEEQKKKAIKWLMDVQEQNGSWYKWKGI CYIYGTWATMTGLRALGVPSHPS  
LKAASWLEHLQYEDGGWGES CQSSVEKKFISLPFSTPSQTAWALDALISYDQETPI IRKGIT  
YLLAQSTMNEKYPTGTGLPGGFYIRYHSYGHYIPLLA LAHYVKKYK

&gt;seq\_ID 64

MSNLLYKVEHEEIARRTTALQTMQRQDGTWRFCEGAPLTDCHMIFLLKLLGRNDEI EPFVKR  
LASLQTNEGTWKLYEDEDGNNLSATIQSYAALLASEKYTKEDANMKRAEMPI INERGGVARAHF  
MTKFLLAIHGEYEPALPHFPPTLFLQDSDPLSIFELSSSARIHLIPMMLCLNKRFRVGGKLLPNL  
NHIAGGGQWFRREERSPLFQSLGDVKKVI IYPLSLHHKGYEEVERFMKERIDENGLTYSYASA  
SFYMIYALLALGHSIQSPIIEKAITGTSYIWKMERGSHLQNSPSTIWDTALLSYALQEQAVPKASK  
VIHNASAYLLRQQT KRMDWSVHAPDLFPGGWGFSVDVNTTIPD IDDTAALRALARSRGNEV  
DNAWKRAVNWVKGLQNNDDGGWGA FERGVTSRILANLPIENASDMI TDPSTPDI TGRVLEFPGT  
YTQNELPEKQKQSAINWLMNVQEENGSWYKWKGI CYIYGTWAVMTGLRFGIPSSNPSLKRA  
ALWLEHIQHEDEGGWGES CQSSVEKRFVTLFPSTPSQTAWALDALISYDQETPI IRKGISYLLS  
NSYINKEYPTGTGLPGGFYIRYHSYAHYIPLLA LAHYAKKYK

-continued

## Enzyme Sequences

&gt;seq\_ID 68

MLLYEKVHEEIIARRTTALQTMQRQDGTWRFCFEGAPLTDCHMIFLLKLLGKDKIEPFVKRLAS  
 LQTNEG TWKLYEDEVGGNLSATI QSYAALLASEKYTKEDANMKRAEMFINERGGVARAHFMTK  
 FLLAIHGVEYEPSLPHLPTPI MFLQNDSPLSIFELSSSARIHLI PMMLCLNKRFRVGGKLLPNLNHI  
 AGGGGEWFPREDRSPVVFQTLVSDVKKII TYPLSLHHKGYEEVERFMKERIDENGLTLYSYATASFY  
 MIYALLALGHSIQSPI IQKAI TGI TSYIWKMERGSHLQNSPSTVWDTALLSYALQEAQVPKASKVI  
 HNASAYLLRKKQTKKVDWSVHAPDLFPGGWGFSVNTTIPDIDDTAALRALARSRGNEVD  
 AWKRAVNWVKGLQNDGGWGAPEKGVTSRILANLPIENASDMI TDPSTPDI TGRVLEFFGTYT  
 QNELPEKQKQSAINWLMNVQEENGSWYKKGWICYI YGTWAVLTGLRSLGIPSSDPSVKRAAL  
 WLEHIQHEDGGWGESCSQSSVEKRFVTL PFSPTSQTAWALDALISYDKEKTPVIRKGISYLLSNS  
 YINEKYPTGTGLPGGFYIRYHSYAHYIPLLLAHLAYAKKYRK

&gt;seq\_ID 41

MSNLLLYEKVHEEIIARRTTALQTMQRQDGTWQFCFEGAPLTDCHMIFLLKLLGRDKIEPFVVKR  
 LASLQTNEGTWKLYEDEVGGNLSATI QSYAALLASEKYTKEDANMKRAEMFINERGGVARAHF  
 MTKFLLAIHGVEYEPSLPHLPTPI MFLQNDSPLSIFELSSSARIHLI PMMLCLNKRFRVGGKLLPNLN  
 NHIAGGGGEWFPREDRSPVVFQTLVSDVKKII TYPLSLHHKGYEEVERFMKERIDENGLTLYSYATA  
 SFYMIYALLALGHSIQSPI IQKAI TGI TSYIWKMERGSHLQNSPSTVWDTALLSYLVQEAQVPKAS  
 KVIHNASAYLLRKKQTKKVDWSVHAPDLFPGGWGFSVNTTIPDIDDTAALRALARSRGNEVD  
 VDTAWKRAVNWVKGLQNDGGWGFPEKGVTSRILANLPIENASDMI TDPSTPDI TGRVLEFFG  
 TYTQNELPEKQKQSAINWLMNVQEENGSWYKKGWICYI YGTWAVLTGLRSLGIPSSDPSVKRA  
 ALWLEHIQHEDGGWGESCSQSSVEKRFVTL PFSPTSQTAWALDALISYDKEKTPVIRKGISYLLS  
 NSYINEKYPTGTGLPGGFYIRYHSYAHYIPLLLAHLAYAKKYRK

&gt;seq\_ID 66

MSNLLLYEKVHEEIIARRTTALQTMQRQDGTWQFCFEGAPLTDCHMIFLLKLLGRDKIEPFVVKR  
 LASLQTNEGTWKLYEDEVGGNLSATI QSYAALLASEKYTKEDANMKRAEMFINERGGVARAHF  
 MTKFLLAIHGVEYEPSLPHLPTPI MFLQNDSPLSIFELSSSARIHLI PMMLCLNKRFRVGGKLLPNLN  
 NHIAGGGGEWFPREDRSPVVFQTLASDVKKII TYPLSLHHKGYEEVERFMKERIDENGLTLYSYATA  
 SFYMIYALLALGHSIQSPI IEKAIMGITSYIWKMERGSHLQNSPSTIWDTALLSYALQEAQVPKAS  
 KVIQNASAYLLRKKQTKKVDWSVHAPDLFPGGWGFSVNTTIPDIDDTAVLRALARSRGNEVD  
 VDNWAKRAVNWVKGLQNDGGWGAPEKGVTSRILANLPIENASDMI TDPSTPDI TGRVLEFFG  
 TYGQNELPEKQKQSAINWLTNAQEENGSWYKKGWICYI YGTWAVLTGLRSLGIPSSDPSVKRA  
 ALWLEHIQHEDGGWGESCSQSSVEKRFVTL PFSPTSQTAWALDALISYDKEKTPVIRKGISYLLS  
 NPYINEKYPTGTGLPGGFYICVHSYAHYIPLLLAHLAYAKKYRK

&gt;seq\_ID 138

MVADERSALIDALKRSQSVDDGSRFPFETGISTDAYMI ILLRTLGIHDEPLIQALVERIESRODAN  
 GAWKLFADEGDGNVTATVEAYALLYSGYRKKTD SHMQAKARILEVGGLEERVHLFTKVMMLAL  
 TGQHSWPRRFPPLPLVFFLLPPSFPPLNMYDLSVYGRANMVP LLVVAERRY SRKTDNSPDLSDLA  
 ASRNDWRLPDTEALWSYVYKRSLTGLPAWLHRAAEQRAVRYMLEHI EPDGTLYSYFSSSTFLLI FA  
 LLALGYPKDDPHIARAVRGLRSLRTEIDGHTMQYTTASVWNTALAS YALQEAQVPPTDRTIEK  
 ANRYLLSRQHIRYGDWAVHNYPVPGGWGFSVNTTIPDIDDTAVLRALARSRGNEVD  
 AWRANRWLFSMQNDGGFAAPEKKNVGRKRFWRYP IEGAEFLMDDPSTADLTGRITLEYFGTF  
 AGLTKDHSARAIARAI DWLDDHQEADGSWYGRWIGICYI YGTWAAVTLGSLAVGVPIDHPAMQKAV  
 RWLLS IQNDGGWGESCKSDGAKTYVPLGASTPVHTAWALDALIAAERPTPEMKAGVRALV  
 RMLHHPDWTASYPVQGMAGAFYIHYHGYRYI PPLALAHAYEQKGPFPVD

&gt;seq\_ID 69

MLLYEKVHEEIIARRTTALQTMQRQDGTWRFCFEGAPLTDCHMIFLLKLLGRDKIEPFVKRLAS  
 LQTNEG TWKLYEDEVGGNLSATI QSYAALLASEKYTKEDANMKRAEMFINERGGVARAHFMTK  
 FLLAVHGEYEPSLPHLPTPI MFLQNDSPLSIFELSSSARIHLI PMMLCLNKRFRVGGKLLPNLNHI  
 AGGGGEWFPREDRSPVVFQTLLEEVKKII TYPLSLHHKGYEAVERFMKERIDENGLTLYSYATASFY  
 MIYALLALGHSIQSPI IQKAI TGI TSYIWKMERGSHLQNSPSTVWDTALLSYALQEAQVPKASKGI  
 QNASAYLLRKKQTKKVDWSVHAPDLFPGGWGFSVNTTIPDIDDTAVLRALARSRGNEVD  
 NSWKRAVNWVKGLQNDGGWGAPEKGVTSRILANLPIENASDMI PDPSTPDI TGRVLEFFGTY  
 AQNELPEKQKQSAINWLMNIQEENGSWYKKGWICYI YGTWAVLTGLRSLGIPSSDPSVKRAAL  
 WLEHIQHEDGGWGESCSQSSVEKRFVTL PFSPTSQTAWALDALISYDKEKTPVIRKGISYLLSNP  
 YVNEKYPTGTGLPGGFYIRYHSYTHIYPLLLAHLAYAKKYRK

&gt;seq\_ID 67

MSNLLLYEKVHEEIIARRTTALQTMQRQDGTWRFCFEGAPLTDCHMIFLLKLLGRDKIEPFVVKR  
 LASLQTNEGTWKLYEDEVGGNLSATI QSYAALLASEKYTKEDANMKRAEMFINERGGVARAHF  
 MTKFLLAIHGVEYEPSLPHLPTPI MFLQNDSPLSIFELSSSARIHLI PMMLCLNKRFRVGGKLLPNLN  
 NHIAGGGGEWFPREDRSPVVFQTLLEEVKKII TYPLSLHHKGYEEVERFMKERIDENGLTLYSYATA  
 SFYMIYALLALGHSIQSPI IQKAI TGI ASYIWKMERGSHLQNSPSTVWDTALLSYALQEAQVPKAS  
 KVIQNASAYLLRKKQTKKVDWSVHAPDLFPGGWGFSVNTTIPDIDDTAVLRALARSRGNEVD  
 VDNWAKRAVNWVKGLQNDGGWGAPEKGVTSRILANLPIENASDMI TDPSTPDI TGRVLEFFG  
 TYAQNELPEKQKQSAINWLMNVQEENGSWYKKGWICYI YGTWAVLTGLRSLGIPSSDPSVKRA  
 ALWLEHIQHEDGGWGESCSQSSVEKRFVTL PFSPTSQTAWALDALISYDKEKTPVIRKGISYLLSN  
 PYNKYPTGTGLPGGFYIRYHSYAHYIPLLLAHLAYAKKYRK

&gt;seq\_ID 35

MSNLLLYEKAHEEIVRRATALQTMQWQDGTWRFCFEGAPLTDCHMIFLLKLLGRDKIEPFVVER  
 VASLQTNEGTWKLYEDEVGGNLSATI QSYAALLASKKYTKEDANMKRAEMFINERGGVARAHF  
 MTKFLLAIHGVEYEPSLPHLPTPI MFLQDDAPPSIFELSSSARIHLI PMMLCLNKRFRVGGKLLPN

-continued

Enzyme Sequences

LNHIAGGGGEWFRDRSPVFTLLSDVKQIISYPLSLHHKGYEIERFMKERIDENGLTLYSATA  
 SFYMIYALLLALGHSIQSSMIQKAIAGITSYIWKMERGNHLQNSPSTVWDTALLSYALQEAQVSK  
 DNKMIQNATAYLLKKQHTKKADWSVHAPALTPGGWGFSDVNTTIPDIDDDTTAVLRALARSRGN  
 KNIDNAWKGGNWIKGLQNNDDGGWGAPEKGVTSKLLAKLPIENASDMITDPSTPDI TGRVLEFF  
 GTYAQNELPEKQIQRAINWLMNVQEENGWSYWKWGI CYIYGTWAVMTGLRSLGIPSSNPSLKR  
 AASWLEHIQHEDGGWGESCHSVEKRFVTLPPSTPSQTAWALDALISYYDTETPAIRKGSYLL  
 LNPYVNERYPGTGLPGAFYIRYHSYAHYIPLLLTAHYLKKYRK

>seq\_ID 43  
 MNALLLYEKVHEEIIARRTTALQTMQRQDGTWRFCEGAPLTDCHMIFLLKLLGRDKEVEPFVK  
 RLASLQTNEGTWKLYEDEVGGNLSATI QSYAALLASKKYTKEDANMKRAEMFITERGGVARAH  
 FMTKFLLAIHGEYEYPSLFHLPPTIMFQNDSPSLIFELSSSARIHLIPMMLCLNKRFRVGGKLLP  
 NLNHIAGGGGEWFRDQSPMPQTLGNVQKII SYPLSLHHKGNEEVERFMKERIDENGLTLYS  
 ASAFYMIYALLLALGHSIQSPMIQKAI TGI TSYIWKMERGNHLQNSPSTVWDTALLSYALQEARV  
 SKESKMIQNASAYLLKKQHTKKADWSVHAPVLI PGGWGFSDVNTTVPDVEDTTAVLRALQSR  
 GNGNVDDAWKGTNWI KGLQNNDDGGWGAPEKGVTSKLLANLPI ENASDMITDPSTPDI TGRVL  
 EFGTYTQNELPEKQKQSAINWLMNVEENGWSYWKWGI CYIYGTWAVMTGLRSLGITSAP  
 SLKRATLWLEHIQHEDGGWGESCSQSSVEKRFATLPPSTPSQTAWALDALISYYDKETPAIRKGI  
 SYLLANPYVNEKYPGTALPGGFYIHYHSYAHYIPLLLTAHYAKKYK

>seq\_ID 33  
 MNIVIRISKGVVSNLLLYEKVHEEIIARRTTALQTMQRQDGTWQFCFEGAPLTDCHMIFLLKLLG  
 RDKIEPPFVKRLASLQTNEGTWKLYEDEVGGNLSATI QSYAALLASKKYTKEDANMKRAEMFIN  
 ERGGVARAHFMTKFLLAIHGEYEYPSLFHLPPTIMFQNDSPSLIFELSSSARIHLIPMMVCLNKR  
 FQVGGKLLPNLNHIAGGGGEWFRDRSPMFQTLSDVKQIISYPLSLHHKGYEEVERFMKERID  
 ENGLTLYSATA SFYMIYALLLALGHSIQSSMIQKAIAGITSYIWKMERGNHLQNSPSTVWDTALLS  
 YTLQEAHASKDNKMIQHAAAYVLLKQHTKKADWSVHAPGLIPGGWGFSDVNTTIPDVEDTTAV  
 LRALARSRGNENVDNAWKGVNWKGLQNNDDGGWGAPEKGVTSNLLANLPI ENASDMITDPS  
 TPDITGRVLELFGTYAQNELPEKQKQSAINWLMNVEENGWSYWKWGI CYIYGTWAVMTGLR  
 SLGIPSSNPSMKRAALWLEHIQHEDGGWGESCSQSSVEKRFITLPPSTPSQTAWALDALISYHDE  
 ETPAIRKGISYLLANPYVNEKYPGTGLPGGFYIHYHSYAHYIPLLLTAHYIKKYK

>seq\_ID 36  
 MSNLLLYEKVHEEIIARRATALQTMQRQDGTWRFCEGAPLTDCHMIFLLKLLGRDKEIEPPVKR  
 LASLQTNEGTWKLYEDEVGGNLSATI QSYAALLASQKYTKEDANMKRAENFITERGGVARAHF  
 MTKFLLAIHGEYEYPSLFHVPPTIMFQNDSPSLIFELSSSARIHLIPMMVCLNKRFRVGGKLLPN  
 LNHIAGGGGEWFRDRSPVFTLLSDVKQIISYPLSLHHKGYEEVERFMKERIDENGLTLYSATA  
 ASFYMIYALLLALGHSIQSSMIQKAIAGITSYIWKMERGSHLQNSPSTVWDTALLSYALQEAQVSK  
 DHKMIQQTITYLKQHTKKADWSVHAPALTPGGWGFSDVNTTIPDVEDTTAVLRVLRAREN  
 EKVNNAWKQKIDWV KGLQNNDDGGWGAPEKGVTSKLLANLPI ENASDMITDPSTPDI TGRVLEL  
 FGTYTQNELPEKQKQSAINWLMNVAQEENGWSYWKWGI CYIYGTWAVMTGLRSLGIPSSNPSL  
 KRAALWLEHIQHEDGGWGESCSQSSMEKRFITLPPSTPSQTAWALDALISYYDTETPAIRKGISY  
 LLANPYVNEKYPGTGLPGGFYIRYHSYAQIYPLLLTAHYTKKYK

>seq\_ID 42  
 MSNLLLYEKVHEEIIARRTTALQTMQRQDGTWRFCEGAPLTDCHMIFLLKLLGRDKEIEPPVKR  
 LASLQTNEGTWKLYEDEVGGNLSATI QSYAALLASEKYTKEDANMKRAEMFINERGGVARAHF  
 MTKFLLAIHGEYEYPSLFHLPPTIMFQNDSPSLIFELSSSARIHLIPMMLCLNKRFRVGGKLLPNL  
 NHIAGGGGEWFRDRSPVFTLVSDVKKII TYPLSLHHKGYEEVERFMKERIDENGLTLYSATA  
 SFYMIYALLLALGHSIQSPI IEKAIMGITSYIWKVERGSHLQNSPSTIWDTALLSYALQEAQVSK  
 VIQNASAYLLRKQQTKKVDWSVHAPDLFPGGWGFSDVNTTIPDIDDDTTAVLRALARSRGNEHV  
 DNAWKRAVNWV KGLQNNDDGGWGAPEKGVTSRI LANLPI ENASDMITDPSTPDI TGRVLEFFGT  
 YTQNELPEKQKQSAINWLMNVAQEENGWSYWKWGI CYIYGTWAVMTGLRSLGIPSSDSSLKRAV  
 LWLEHIQHEDGGWGESCSQSSVEKRFVTLPPSTPSQTAWALDALISYYDKETPVIRKGISYLLSN  
 PYINEKYPGTGLPGGFYIRYHSYAHYIPLLLTAHYAKKYK

>seq\_ID 65  
 MSNLLLYEKVHEEIIARRTTALQTMQRQDGTWRFCEGAPLTDCHMIFLLKLLGRDKEIEPPVKR  
 LASLQTNEGTWKLYEDEVGGNLSATI QSYAALLASEKYTKEDANMKRAEMFINERGGVARAHF  
 MTKFLLAIHGEYEYPSLFHLPPTIMFQNDSPSLIFELSSSARIHLIPMMLCLNKRFRVGGKLLPNL  
 NHIAGGGGEWFRDRSPVFTLVSDVKKII TYPLSLHHKGYEEVERFMKERIDENGLTLYSATA  
 SFYMIYALLLALGHSIQSPI IEKAIMGITSYIWKMERGSHLQNSPSTIWDTALLSYALQEAQVSK  
 KVIQNASAYLLRKQQTKKVDWSVHAPDLFPGGWGFSDVNTTIPDIDDDTTAVLRALARSRGNEV  
 VDNAWKRAVNWV KGLQNNDDGGWGAPEKGVTSRI LANLPI ENASDMITDPSTPDI TGRVLEFFGT  
 TYTQNELPEKQKQSAINWLMNVAQEENGWSYWKWGI CYIYGTWAVMTGLRSLGIPSSDSSLKRAV  
 VLWLEHIQHEDGGWGESCSQSSVEKRFVTLPPSTPSQTAWALDALISYYDKETPVIRKGISYLLS  
 NPYINEKYPGTGLPGGFYIRYHSYAHYIPLLLTAHYAKKYK

>seq\_ID 39  
 MNLLLYEKVHEEIIARRATALQTMQRQDGTWRFCEGAPLTDCHMIFLLKLLGRDKEIEPPVKR  
 LASLQTNEGTWKLYEDEVGGNLSATI QSYAALLASKKYTKEDANMKRAENFITERGGVARAHF  
 MTKFLLAIHGEYEYPSLFHLPPTIMFQNDSPSLIFELSSSARIHLIPMMLCLNKRFRVGGKLLP  
 NHIAGGGGEWFRDRSPVFTLVSDVKQIISYPLSLHHKGYEEVERFMKERIDENGLTLYSATA  
 SFYMIYALLLALGHSIQSPTMIQKAI TGI TSYIWKMESGNHLQNSPSTVWDTALLSYALQEAHVPKD  
 NKMIQHAATYLLKKQHTQKADWSVHAPALTPGGWGFSDVNTTIPDVEDTTAVLRALARSRGNE  
 KVDNAWPKGINWV KGLQNNDDGGWGAPEKGVTSNLANLPI ENASDMITDPSTPDI TGRVLEFF  
 GKYAQNELPEKQKQSAINWLMNVAQEENGWSYWKWGI CYIYGTWAVMTGLRSLGIPSSNPSMK

-continued

## Enzyme Sequences

RAALWLEHIQHEDGGWGESCHSSVEKRFVTLFPSTPSQTAWALDALISYYDKETSIRKGISYLL  
ANPYVNEKYPTGTGLPGGFYIRYHSYAHYIPLLLTLAHYIKKYRK

>seq\_ID 63

MSNLLLYEKAHEEIIARRATALQTMQREDGTWRFCFEGAPLTDCHMIFLLKLLGRDKEIEPFFVKR  
LATLQTNEGTWKLYEDEVGGNLSATIQSYAALLASGKYTKEDANMKRAENFIKERGGVARAHF  
MTKFLLAIHGEYEPYPSLFHVPPTIMFLQNDSPSIFELSSSARIHLIPMMLCLNKRFRVGGKLLPN  
LNHIAGGGGEWFREREPLFQTLSDVKQIISYPLSLHHKGYEEVERFMKERIDENGLTLYSYAT  
ASFYMIYALLALGHSLOQSMIQKAIAGITSIYWKMESGNHVQNSPSTVWDTALLSYALQEAHVP  
KDNKMLQNAATAYLLKKQHTKKADWSVHAPALTPGGWGFSDVNTTVPDVTAVLRVLRARSK  
GNEKLDHAWQKGINWVKGLQNNDDGGWGAFAKGVTSRILANLPIENASDMITDPSTPDI TGRVLE  
EFFGTYAQNELPEKQKQSAINWLMNAQEENGSWYKKGWICYIYGTWAVMTGLRSPGIPSSNP  
SLKRAALWLEHIQHEDGGWGESCHSSVEKRFVTLFPSTPSQTAWALDALISYYDTETPVIKGI  
SYLLANPYVNEKYPTGTGLPGGFYIRYHSYAHYIPLLLTLTHYIKNIENKPRDISRIFLGRSLLKRI  
RLCFPPYFSDWRF

>seq\_ID 37

MSNLLLYEKAHEEIIARRATALQTMQREDGTWRFCFEGAPLTDCHMIFLLKLLGRDKEIEPFFVKR  
LASLQTNEGTWKLYEDEVGGNLSATIQSYAALLASGKYTKEDANMKRAENFIKERGGVARAHF  
MTKFLLAIHGEYEPYPSLFHVPPTIMFLQNDSPSIFELSSSARIHLIPMMLCLNKRFRVGGKLLPN  
LNHIAGGGGEWFREREPLFQTLSDVKQIISYPLSLHHKGYEEVERFMKERIDENGLTLYSYAT  
ASFYMIYALLALGHSLOQSMIQKAIAGITSIYWKMESGNHVQNSPSTVWDTALLSYALQEAHVP  
KDNKMLQNAATAYLLKKQHTKKADWSVHAPALTPGGWGFSDVNTTVPDVTAVLRVLRARSK  
GNEKLDHAWQKGINWVKGLQNNDDGGWGAFAKGVTSRILANLPIENASDMITDPSTPDI TGRVLE  
EFFGTYAQNELPEKQKQSAINWLMNAQEENGSWYKKGWICYIYGTWAVMTGLRSPGIPSSNP  
SLKRAALWLEHIQHEDGGWGESCHSSVEKRFVTLFPSTPSQTAWALDALISYYDTETPVIKGI  
SYLLANPYVNEKYPTGTGLPGGFYIRYHSYAHYIPLLLTLTHYIKKYRK

>seq\_ID 46

MLLYEKVHEEVKEKMAALQAMQQDGTWRFCFEGSPLTDCYMIPLLLGGQDQIEPFFVARLA  
ALQTNEGTWKLYEDEPDGNLSATIQAAYALLVSKMYKKEDINMKRAEVIRKQGGITKAHFMTK  
FLLALHGGYEPYPLFHPPTPIFLFSEDSPLSIFELSSSARIHLIPMMLCMNKRFTVSKKMLPNLDYI  
SGGSKEQWFREREPLFQTLRDVTKFLSYPLSLHYKGDKAERFMIERIDENGLTLYSYASATF  
YMIYALLALGHSIQSPLISNAVLGLKTYVWVMDRWAHLQNSPSTVWDTALLSYLQEARVPHD  
NEMIQKAINYLLQKQKQKADWSVHAPALDAGGWFSDVNTTIPDVTAVLRALAGSRQGN  
PKVESAWRKGLEWVKGLQNSDGGWAAFAKGVTSKVLTHLPLDNGSDMITDPSTVDITGRVLEF  
FGTYAPNELQGDQKDRIRWLIYTOEKNGSWHGKGVVCIYGTWAAALGLRAVGVPSNHIAL  
QKAATWLESIQHSDGGWGESCRSSVEKRFVTLFPSTPSQTAWALDALIACYDSETPIRKGISYLL  
LKHS TKHQEYPTGTALANGFYIRYHSYHHIPLLLTFAHYIKKYRK

>seq\_ID 40

MSNLLLYEKVHEEIIARRATLQTMQRDGTWRFCFEGAPLTDCHMIFLLKLLGRDKEIEPFFVKR  
LASLQTNEGTWKLYEDEVGGNLSATIQSYAALLASGKYTKEDANMKRAENFINERGGVARAHF  
MTKFLLAHGEYEPYPSLFHLPPTIMFLQSDSPSIFELSSSARIHLIPMMLCLNKKFRIRKLLPNL  
NHSIGGGGEWFRERNRSPFLQTLVSDVKQIISYPLSLHHKGYEEVERFMKERIDENGLTLYSYATA  
SFYMIYALLALGHSIQSPLIQKAIAGITSIYWKMESGNHLQNSPSTVWDTALLSYALQEAHVPKD  
TNMLQHATAYLLKKQHTKKADWSVHAPALAPGGWGFSDVNTTIPDVTAVLRALARSRG  
EKVDYVWEKGINWVKGLQNNDDGGWGAFAKGVTSNLLANLPIENASDMITDPSTPDI TGRVLEL  
FGTYAQNELPEKQKQSAINWLMNVQEKNGSWYKKGWICYIYGTWAVMTGLRSLGIPSSNP  
KRAALWLEHIQHEDGGWGESCHSSVEKRFVTLFPSTPSQTAWALDALISYYDKETPAIRKGISY  
LLANRYVNEKYPTGTGLPGGFYIRYHSYAHYIPLLLTLAHYIKKYRK

>seq\_ID 38

MSNLLLYEKAHEEIIARRATALQSMQWQDGTWRFCFEGAPLTDCHMIFLLKLLGRDKEIEPFFVK  
RLASLQTNEGTWKLYEDEVGGNLSATIQSYAALLASGKYTKEDANMKRAENFIKERGGVARAH  
FMTKFLLAHGEYEPYPSLFHLPPTIMFLQNDSPSIFELSSSARIHLIPMMLCLNKRFRVGGKLLPN  
LNHIAGGGGEWFREREPLFQTLVSDVKQIISYPLSLHHKGYEEVERFMKERIDENGLTLYSYA  
TASFYMIYALLALGHSIQSPLIQNAITGITSYWKMESGNHLQNSPSTVWDTALLSYALQEAHVPK  
DNKMLQNAATAYLLKKQHTKKADWSVHASALTPGGWGFSDVNTTIPDVTAVLRVLRARSRG  
NEKVDHAWQKGINWVKGLQNNDDGGWGAFAKGVTSNLLAKLPIENASDMITDPSTPDI TGRVLE  
FFGTYAQNELPEKQKQSAINWLMNVQEKNGSWYKKGWICYIYGTWAVMTGLRSPGIPSSNP  
LKRAALWLEHIQHEDGGWGESCHSSVEKRFVTLFPSTPSQTAWALDALISYYDTETPIIRKGISY  
LLANRYVNEKYPTGTGLPGGFYIRYHSYAHYIPLLLTLAHYIKKYRK

>seq\_ID 55

MLLYEKVQRQEVERKVTALRTMQYQDGAWRFCFEGSPLTDCYMIPLLLGGQNGEMEPFVTRV  
ASLQTNEGTWKLYEDES VGNLSTTINAVALASGRYTKEDINMKRAEAFIRRQGGITKAHFMT  
KFLALHGGYEPYPSLFHLPPTMFLPLPDSPLSIFELSSSARIHLIPMMI CMNKRFTVSKITLPLNLDY  
ISGGSKQWFRERESSLQRLLGDVKKFLSYPLSLQHKGYKAEERFMIERIDENGLTLYSYASAT  
FYMIYALLALGHSIQSPLISNAVLGLKSYIWNMNGKTHLQNSPSTVWDTALLSYLQEAHVND  
NQMIQKATDYLLQKQKQKADWSVHAPSLDAGGWFSDVNTTIPDIDDTAALRAIARSREGN  
QRIEEAWRKGLEWVKGLQNDGGWAAFERGVTSHFLTHLPLDNGDMMITDPSTSDITGRVLEF  
FGTYAPHQLKDDQKDRIKWLMQAQEKNGSWYKKGWICYIYGTWAAALGLRAVGVPSNHATA  
LQKAATWLERIQHNDGGWGESCRSSIEKHFI SLFPSTPSQTAWALDALITFYDTETPVIKGISY  
LLAHLNQNQDYPTGTGLPDGFYIRYHSYHHIPLLLTFAHYIKKYMK

-continued

## Enzyme Sequences

&gt;seq\_ID 54

MLLYEKVRQEVERKVTALRRTTQYQDGAWRFCFEGSPLTDCHMI FLRLRLGQNGEMEPFVTRV  
 ASLQTNEGTWKLYEDES VGNLSTTINAYVALLASGRYTKEDINMKRAEAFIRROGGITKAHFMT  
 KFLALHGGYEPYPSLFHPTPLFLPEDSPLSIFELSSSARIHLIPMMI CMNKRFTVSKTIFPNLDY  
 ISGGSKQWFREERSPLFQTLGDDVKKPLSYPLSLQHKGYKEAERFMIERI ETNGTLYSYASAT  
 FYMIYALLALGHSIQSPLI SNAVLGLKSYIWNMKGTHLQNSPSTVWDTALLSYSLQEAGVPND  
 NQMIQKATDYLLQKQHKKEDWSVHAPSLDAGGWGFSVDVNTTIPDIDDTAALRAIARSREGN  
 QRIEEDWRKGI EWFVKLQNLIDGGWAAFERGVTSHFLTHLPLDNAGDMTDPSTSDITGRVLEF  
 FGTYAPHQLKDDQKDRAIKWLMOAQEKNGSWYKGWVCYI YGTWAVLTGLRAVGVPSNHTA  
 LQKAATWLERIQHNDGGWGESCRSSIEKHFI SLPFPSTPQTAWALDALI TPYDTETPVIRKGISY  
 LLAHLNQNQDYPTGIGLDPDGFYIRYHSYHHIFPILTFAHYIKKYMK

&gt;seq\_ID 189

MRSELLQLQSDAGSWRLCFDSGTMPSYFII LRMLGYSQDEALIRQIASRILSRQLPNGTWKI Y  
 PDEEDGNLDATAEAYFALLYSGFLTKLDPRMQLAKQFI LSKGGLSKIRSLLTQAI FAAAAGQASWP  
 KSMRIPLEVFFSDNGIGIDFLSLSGHARVHIVPI IMLANAQFVQHSASMPDLSDLFAGSSKRFEN  
 DSPWIAALATLIGSLSELSELPFESPTPQEKAVQFLFDRLPEPDGTLTYTTATMFMILVLLMLGYS  
 SSSPLIHRMVSIGHSIVI CANSHVQIASSEVVD TAMLVHALRKAGVNPSTALENAGAYLRQRQ  
 TQLGDWAIARNPGTPAGGWGFSNVNTLYPDVDDTTAALRAIQPYSSRTPELQADWQRLNWL  
 TMRNDNGGWPAFERQGSRLPI TFFNFEGAKDIAVDPSTVDLTSRTLQPLGQELGMNAGNSWIE  
 STLRWVLSQQESNGSWYGRWGI TYVHGTSAAQLGLTAVGIAEDHPAVKKGVDWLLQVQNE  
 GGWGESCI SDKVRRYVPLNFS TSPQTAWALDGLTAAALPKPTPALERGDALLQSLDRHDWTY  
 TYPTGGALPGSVYAHYASNNYIWPLLALSNIWQKYS

&gt;seq\_ID 200

MALPFNQDSYKGDDEADVSKGAAKSPSLEEAIQRSQEFLLAQQFPEGFWFGELEANVTI ISHT  
 VILYKLLGIEENFPMYKPERYLRMQCSHGGWEIAYGIGSYLSATIEAYIALRLLNVPQSDPALQK  
 ALRVILDSGGVTKARIFTKICLALLGSPDWRGIPSLPPWILCPTWPLSI YEVSWSARGCIVPLL  
 VLLDKKPVFKVSPVYSFDELAYEGREHACKI IPI SGDWTSKFFITVDRVFKMMERLRVVPFRQW  
 GIREAEKWILERQEBESGDYVNI FPPAMFYVVMCMKVLGYETTPDVVQRALLGFKGFTIETADECK  
 VQSTVSPITWDTAFI VRALVDSGI PPDHPALQKAGQWLLQKQILKHGDWAFKDRQNPVNRGFA  
 CLQRDSQIETADECRVQSTLS PVWD TAFVVKALVDSGI PPNHPALQKAGQWLLQNTLTHGD  
 WAFKTSQSHLAAGGWAFQSHNRWYDADDSAAVMALDCI ELPEDEVKNGAIARGLKWI SAL  
 QSRNGGAGYDKNCQQWINKVFPNDLNGI LDVPTADV TARVLEMVGRLSRLGAVGTPYSPR  
 HCTLVESI PHLLLPETIARGLAYLRREQEGEGCWGKGVNYI YGTCGALLALSQVAPTTHQE  
 ETARGAKWLAQVQRCDKQAAQGRDGGWGESCF SYDDPALKGQNDASTASQTAWAVQG  
 LLAAGDALGKYEVEAIEQGVQYLLATQRKDG TWHEAHFTGSCFAQH FVRYHYAQHFPLSAL  
 GLYRTRILQHQ

&gt;seq\_ID 139

MVADERSALIDALKRSQSVGDSWRFPFETGISTDAYMI ILLRTLGIHDEPLIQALVERIESRQDAN  
 GAWKLFADDEGDGNVTATVEAYYALLSYGYRKKTD SHMQAKAKARILEVGGLEVRVHLFTKVMLAL  
 TGQHSWPRRFPPLPVFPLPPSPFLNMVYDLSVYGRANMVP LLVVAERRY SRKTDNSPDLSDLA  
 ASRNDWRLPDTEALWSYVKRSLTGLPAWLHRAAEQRAVRYMLEHI EPDGTLYSYFSS TFLLI FA  
 LLALGYPKDDPHIARAVRGLRSLRTEIDGHTHMQYTTASVWNTALAS YALQEAGVPPTDRTIEK  
 ANRYLLSRQHIRYGDWAVHNPYGVPGWGFSDVNTMNPVDDTTAALRAIRRAAAKETAFRH  
 AWRANRWLFSM QNDGGFAAEKKNVGRFWR YLPI EGAEFLMDPS TADLTGR TLEYFGTF  
 AGLTKDHSIARAIADWLLDHQEDAGDSWYGRWGCYVYGTWAAVTGLS AVGVPIDHPAMQKAV  
 RWLLSI QNDGGWGESCKSDGAKTYVPLGASTPVHTAWALDALIAAERPTPEMKAGVRAV  
 RMLHHPDWTASYPVGGMAGAFYIHYHG YRYI FPLLALAHYEQKFGPFD

&gt;seq\_ID 13

MAQMASLGS PRLLLRMGREAQQQHLASGTEVQKALRLAVGHS LLDLQRTDGAWCGEVHSN  
 ATFTAQYVFLQQQIGLPLDPTIEGLSRWLF SQQNEGDSWGLGPGLDGVS TTTETYLALKILG  
 VSPEDPRMAAARTS I IKAGSLPATRMFTRVPLAS FGLI PWSAVPPLPAELI LLLPTLFPVNIYNLSS  
 WARA TCVPLLLIRHHEPLHSLPNGRHAENDFLDELWTKDI PRDFCYTTPLSRMMWRLGDYAGIFF  
 TSADHGFRFLGQYFN SPLRNL SRRKI INWI LDHQEQSGEWAGYVVPQHNNI WALSLEGYSLDH  
 PVLRRGIAAVKSFV LHDATGMRAQVTVSQVWD TALMSI TALS DSAPSTGI ISPTQAI DWLMHHEV  
 ASHRGDWRVLRK LATGGFCFEFN TLYPDVDDTAAVIMALI KSNPAHLISGCVRQCFGMMMA  
 GRHGYS LDCQLETRLRASSQLAIAYLLGCQENNGS WWRGWVNYLYGTSNVLCGLAYYDR  
 SLSKGDGKSNNSNI VSAVDRASEWLKARQHSNGGWEGLESYD NAQLAGCGOPTASQSAW  
 VTMALLNYLSPTDEVIQRGVS YLVRNQVKYGDSE RATWPLERYTATGFPGLHYMEYDYRHYF  
 PIMALGRYVKNLGS SHKLL

&gt;seq\_ID 198

MEDLTQKLQALQALASRALNERNRPLGLAHWEGELSTSALSTATAVMALFQYAKCQQASGRL  
 QKVFQKSEGEWRLIEQGLAWLLQHQLADGGWGD TDKISNISTTMLAHATLVACREAVRQK  
 SLVNLASDIDAAIERSGRLEELGGIQAIRDYRQKDTFSVPILTHAALAGLVSWNEIPALPYELAL  
 LPHRFVEVIQLPVVSYALPALIAIGQTLHLRQRTWNPWWVRRRAIPGTLQKLQSIQPESEGGFL  
 EATPLTSPVMTCLASVGRVDHPVTQAGLKFIRDSV RPDGSWPIDTNLATWVTTLS INHLGAEAF  
 SDDEREA LMRWLLQQQYRTMHFY TNAAPGGWAWNTL SGGV PDADDTPGAMLALMELDRVS  
 VSSQESLSIEQALYQAALWLIKLNDRGGWPTFCRGGALPFDPSNDI TAHLRALI QYERRL  
 NDVTVDATGDTT SRPLAVEVSPKLRQMQRS IQQGFEYLEKTQREDGWSLPLWFGNQHS PD  
 DENPLYGTARVLLAYADAGLEGSSAALRGCDWLVRHQHADGAWGPGTSETADTSDAESDVE  
 GEPASIEETALALMALCRFDATHNVLRGASWLI TKVENETWREPPIGFIYAKLWYIEKLYPQ  
 VFTVGALKALALRLGSALT TVSENEPAPSSAEPP IPPIATDRVADSMHLQRTSPSINLANGGITLA

-continued

## Enzyme Sequences

>seq\_ID 252  
 SPVWDTVLTLLALDDCCGYNDYCYSEEVDKAVQWVLDQOVLSKGDWSVKLPNVEPGGWAFEYA  
 NTRYPD TDDTAVALIVLSQFKDDPKWKERGINQAI ERGVNWL FEMQC KNGGWA FDKDNDKT  
 LLTKIPFCDFGEALDPPSVDVTAHIVEAFGLKGLSKDHPKIAHAI EYLKKEEQEADGAWFGRWGV  
 NYVYGTGAVLPALEAIGEDMSQPYIRKAAANWLVLHQ NEDGGWGE

>seq\_ID 253  
 SPVWDTVLTLLAFDDCCDNEAYQASVEKAVQWVLDNQVLRKGDWSVKLPDVEPGGWAFEYA  
 NTFYPD TDDTAVALIVLSQFRDVEKQWEAGIEKAI ERGVNWL FAMSQ KNGGWA FDKDNDNN  
 FITKIPFCDFGEALDPPSVDVTAHCIEAFGLKGLSRARPEIARGLDYLSKSEQEQEADGAWFGRWGV  
 NYVYGTGAVLPALEAIGEDMSQPYIRKAAANWVILRQ NEDGGWGE

>seq\_ID 257  
 SPVWDTXLTLLALDDCDLNERQSKVEKAVQWVLDNQVLRPGDWCVKVPKVPQGGWAFEYK  
 NYFYPD TDDTAVALIVLSQFRDDPKWQEKNI EQAIDRGLNWLIGMQCKGGGWA FDKDNDKT  
 YLTKIPFCDFGEALDPPSVDVTAHIVEAFGLKGLSKDHPNIRRAIDYLSKAEQEQDQAWFGRWGV  
 NYIYGTGAVLPALEAIGEDMRAPYIAKACDWLIVAVQ QEDGGWGE

>seq\_ID 254  
 SPVWDTLLTLLAYDDSGQNERKADEVKAVDQVLAQVLRPGDWKVKAPNLEPGGWAFEYA  
 NYFYPD TDDTAVALIVLSQFRNDAAWKEKIEQAIEKGVNWLFGMQCKGGGWA FDKDNDKQ  
 FLTKIPFCDFGEALDPPSVDVTAHIVEAFGLKGLSKDHPNIRRAIDYMKDEQEQEADGAWFGRWGV  
 NYIYGTGAVLPALEAIGEDMFAPCIGRACDWLVSRQ NDDGGWGE

>seq\_ID 255  
 SPVWDTLLTLLAYDNSGHNARKASEVEKAVDQVLAQVLRPGDWNVKAAPNLEPGGWAFEYA  
 NYFYPD TDDTAVALIVLSQFRNDAAWKDKGIEQAIEKGVNWLFGMQCKGGGWA FDKDNDNR  
 QFLTKIPFCDFGEALDPPSVDVTAHIVEAFGLKGLSKDHPNIRRAIDYTKDEQEQEDDQAWFGRWGV  
 VNYIYGTGAVLPALEAIGEDMSAPYIGRACDWLVSRQ NDDGGWGE

>seq\_ID 256  
 SPVWDTLLTLLAIEDSGQSVKRAQEVKAVDQVLSQVLRPGDWKVRAPHLEPGGWAFEYAN  
 YFFPD TDDTAVALIVLSQFRNDAAWKAKGIEQAI EKGVNWL LGMQCKGGGWA FDKDNDKTYL  
 TKIPFCDFGEALDPPSVDVTAHIVEAFGLKGLSKDHPNIRRAIEYLKSEBQESDQXWFGRWGVNY  
 VYGVGAVLPALEAIGEDMSAPYIGRACDWLVSKQNSDGGWGE

>seq\_ID 258  
 SPVWDTVLTMLAIHDCGADKQYAPQMDKAI DWLLANEVVRHKGDWAVKLPDVEPGGWAFEYS  
 NACYPD LDDTAVALIVLAPYRNDPKWQARDIEGAVERA VDWTLAMQCKNGGWA FDKDNDK  
 AILTKIPFCDFGEALDPPSVDVTAHVLEALALG YDNSHPAVARAIRYL RDEQEPDGSWWGRV  
 GVNYYGTAAVLPALKAMGVMNNEP FVHKAADWIGSVQ NEDGGWGE

>seq\_ID 302  
 SPVWDTSLVLVAMQEAGVPVDHPALVKAQWLLDREVR LKGDWRVKS PDLEPGGWAFEF LN  
 DWYPD VDDSGFVMLALKDI KVRDKKQKSQA I KRG IAWCLGMQ SANGGWA FDKDNTKYLLNK  
 I PFADLEALIDPPTADLTGRMLELMGTFNYPKSHVAVVRLALGLKSVQEP EPGPWWGRWGVNY  
 YGTWQVLRGLAAI DEDMSQPYIRKAVNWLKSKQNL DGGWGEVCE TYEDRS LMGC GPSTPSQ  
 TSWALLSLFSA GEINAKAVLRGI KYLVETQ NQDGSWDE DAYTGTGFP

>seq\_ID 271  
 SPVWDTAISVISLAXSGMERGHPALVRAAXWLMSKEIKTAGDWKVTNPAGV VGGWAFEFNNA  
 FYPD IDDSAMVMALRHVHLDEHTAHRREKACLRLGNLWLSMQSRTGGWA AFDKDNTKVI MT  
 KIPFDLALIDPPTADLTGRMLELMGTFNYPKSHVAVVRLALGLKSVQEP EPGPWWGRWGVNY  
 IYGTWQVLRGLAAI DEDMSQPYIRRAAEWLRSVQPPDGGWGETCATYHDP SLKKG GPATPAQ  
 TAWAVMGLMAAGIYDES VSRGIDYLVLR TQRPDGTWDETEY TGTGFP

>seq\_ID 299  
 SPVWDTALVVLVAMQEAGVPVDHPALIKSAQWLLDLEVR RKGDWHVKS PDLEPGGWAFESLND  
 WYPD VDDSGFVMLFKDI KVRDKKLDQAI KCGIAWCLGMQSENGGWA FDKDNTKHLNKKIP  
 FADLEALIDPPTADLTGRMLELMGTFNYPKSHQAAVKALDFLKVEQEP EPGPWWGRWGVNY  
 GTWSVLCGLEAIGEDMSQPYIKAVNWLKSKQNL DGGWGEVCD SYADRS LMGC GPSTASQT  
 SWALLSLFAAGEVSKAALRGVEYLLSTQKLDGTWDEDAFTGTGFP

>seq\_ID 314  
 SPVWDTALAVRALAAGVPPPEHPAMVKASEWLLTQQIFKPGDWSIKCPDLPPGGWAFEFVNN  
 WYPD VDDSSMVLVALKDGGLADA AKHQALQRGINWCLGMQSKNGGFASFDKDNTKEWLNSL  
 PFGDLKALVDPPTADLTGRMLELMGTFNYPKSHQAAVKALDFLKVEQEP EPGPWWGRWGVNY  
 YGTWQVLRGLAAI DEDMSQPYIRRAAEWLRSVQPPDGGWGETCATYHDP SLKKG GPATPAQ  
 QTAWALLGLFAAGEVHAEVTAAGVYLVKTQD SLGRWDEEQFTGTGFP

>seq\_ID 251  
 SPVWDTVLTMLSVQDCDADENSENAPAI EKAI EWLLANEVRTGGDWQEKVKVEPGGWAFEY  
 KNASYPD TDDTAVALIVLSQFRDVEKQWEAGIEKAI ERGVNWL FEMQC KNGGWA FDKDND  
 KTLCKIPFCDFGEALDPPSVDVTAHVLEGLAALDYPPPEHPAI QRAVQIKDEQEPDGSWWGR  
 WGVNFIYGTAAALPALKAVGEDMRAPYIDRAAKWIVDHQ NEDGGWGE

-continued

## Enzyme Sequences

>seq\_ID 312  
 SPVWDTALAVRALAAGVPPEHPAMVQASEWLLTQQIFKPGDWSVKCPDLP PGGWAFEFVN  
 NWYPDVDDSSMVLVALKDGGLADAQKHAALQRGINWCLGMQSKNGGFASFDKDNTEWLN  
 I PFGDLKALVDPPTEDI TARI LEMMGAFGHGLDHPVAVRAMAYLHETORPEGPWWGRVNYI  
 YGTWSVLVALKRIGEDMSRPYVRAVDWVKAHQNLDDGGWGECCESYRNPPELMGRGPSTAS  
 QTAWALLGLFASGEVHTPEVKAGVDYLVKTKQNSLGRWDEEQFTGTGFP

>seq\_ID 250  
 SPMWDTVLTTLAVQDAGVDQEPEFKPAMERTLEWLLKNEVRTGGDQQKTRGVPEGGWAF  
 EYANASYPDNDDTAVAILVLAFFRHDPKWQARGIQHVIDRAVNWFMAMQCDNNGWAAFDDLN  
 DKAILTRI PFCDFGEALDPPSVDVTAHVLEALALGYSREHPAVRRAIAFLKEDQEPDGSWFGR  
 WGVNFIYGTAAALPALKAMDEDMTQDWITRAADWMSRQNDGGWGE

>seq\_ID 260  
 SPVWDTVLTLLAIQDADKQDDMAAEVDRAIGWLLSKEVRTNGDWSVKLPDVEPGGWAFEHEN  
 ARYPDIDD TAVAVMVLAPYRHHPKWRKRGLPEALDRAI SWMRAMQCRNGGWAFDKDNDN  
 AFLCVI PFCDFGEALDPPSVDVTAHVLEALALGYSREHPAVRRAIAFLKEDQEPDGSWFGR  
 WGVNFIYGTAAALPAYKAFQDMRDPKLMKAADYLRAKQNDGGWGE

>seq\_ID 259  
 SPVWDTVLTLLAMEDCEATEEHAHAIEQAI EWLLNEVTRPGDWQKMPDADPGGWAF EYAN  
 AAYPDVDDTAVAILVLAFFRHDPKWQARGIQHVIDRAVNWFMAMQCDNNGWAAFDDNDKSI  
 LCKI PFCDFGEALDPPSVDVTAHVLEALALGYSREHPAVRRAIAFLKEDQEPDGSWFGR  
 VNYVYGTGAALPAFKAI GADMRDPRMLKAADWILRCQNKDGGWGE

>seq\_ID 261  
 SPVWDTVLTLLAIQDADKQEBMAGEIDKAI GWLLSKEVRTKGDWSVKLPRVEPGGWAF EHEN  
 RYPDIDD TAVAIMVLAPYRHHPKWRKRGLPEALDRAI SWMRAMQCRNGGWAFDKDNDKQIL  
 CTI PFCDFGEALDPPSVDVTAHVLEALALGYSREHPAVRRAIAFLKEDQEPDGSWFGR  
 NYIYGTAAALPAYKALGQDMRDPKLMKAADYLRLDQNDGGWGE

>seq\_ID 262  
 SPVWDTVLTLLAMQDADRDKKHAADV KAIQVWLDQEVTRPGDWCVQTPDVEPGGWAF EYE  
 NARYPDVDDTAVAIMVLAPYRHHPKWRKRGLPEALDRAI SWMRAMQCRNGGWAFDRDNDN  
 SMLTVI PFCDFGEALDPPSVDVTAHVLEALALGYSREHPAVRRAIAFLKEDQEPDGSWFGR  
 WGVNFIYGTSAALPALKAMGRDMDRPRYTKAADYLRAVQNDGGWGE

>seq\_ID 275  
 SPVWDTLLALLALQDCDRELTAEMSRALDWVLANEVRYHGDWTKKVKGVPEPSGWAFERANL  
 NYPDIDD TAVAILVLAFFRHDPKWQARGIQHVIDRAVNWFMAMQCDNNGWAAFDDNDKQIL  
 FCDPGEALDPPSADVTAHVLEALALGYSREHPAVRRAIAFLKEDQEPDGSWFGRVNYVY  
 GTAAVLPGLAAIGEDMTQDYIRRANDWLIHQNDGGWGE

>seq\_ID 280  
 SPVWDTLLSLVALQDCGKELTPARERALEWILGREIRTRGDWAKKVKVNEASGWAFERANLHY  
 PDIDD TAVAILMVLAPYRHHPKWRKRGLPEALDRAI SWMRAMQCRNGGWAFDKDNDKQIL  
 DFCDFGEALDPPSADVTAHVLEALALGYSREHPAVRRAIAFLKEDQEPDGSWFGRVNYVY  
 GAVLPALAAIGEDMAQDYVRRADWLVLAHQNDGGWGE

>seq\_ID 277  
 SPVWDTLLALLAMQDCERELTPOMERALDWVLANEVRYHGDWTKKVKGVPEPSGWAFERANL  
 NYPDIDD TAVAILVLAFFRHDPKWQARGIQHVIDRAVNWFMAMQCDNNGWAAFDDNDKQIL  
 CGFGEALDPPSADVTAHVLEALALGYSREHPAVRRAIAFLKEDQEPDGSWFGRVNYVY  
 GTAAVLPALAAIGEDMSQPYIRAAAEWIIAHQNDGGWGE

>seq\_ID 300  
 SPVWDTALVVLAMQXAGVPVXHPALVKSQWLLDLEVXXKGDWQVKSPELEPGGWAFXFLN  
 DWYPDVDDSGFVMSIKKIKVRDKKHEQAIKRGISWCLGMQSDNNGWAAFDDNDKQIL  
 PFAXLEALIDPPTAXLTGRMLELMGNFNYPKTHKAAVQALEFLXMEPEXPXGFWGRVNYVY  
 GTWSVLCGLEAIGEDMAQPYIKKINWLSKQNDGGWGEVCEVYGRSLMCGGPSTASQT  
 SWALLSLFAAGEVHSAATRGLIEYLLATQKLDGTWEDDAYTGTGFP

>seq\_ID 279  
 SPVWDTLLXLLAMQDCERESTPSMERALDWXXANEVRYHGDWTKKVKGVPEPSGWAFXRANL  
 NYPDIDD TAVAILVLAFFRHDPKWQARGIQHVIDRAVNWFMAMQCDNNGWAAFDDNDKQIL  
 FCDPXEALDPPSADVTAHVLEALALGYSREHPAVRRAIAFLKEDQEPDGSWFGRVNYVY  
 GTXAVLPALAAIGEDMTQPYIRSAAEWIIAHQNDGGWGE

>seq\_ID 264  
 SPVWDTLLTLEALLDNLSPKFTFTGMQAAVDWILSKQIVTPGDWQIKVPGVSCGGWAFERANT  
 FYPMDDTAVAMIVLARIRRYNDSSRIDRALACATDWILSMQCSNNGWAAFDDNDKQIL  
 PFDGEMLDPPSVDVTAHVLEALALGYSREHPAVRRAIAFLKEDQEPDGSWFGRVNYVY  
 YGTGAVLPALAAVGTDMASYITRAADWVATHQNDGGWGE

>seq\_ID 19  
 GGWMFQASISPIWDTGLTVLALRSAGLPPDHPALIKAGEWLVSQKILKDGDKWVRRRKAAPGG  
 WAFEFHCENYPDVDDTAMVLAALNGIQLPDEGKRDALTRGFRWLREMQSSNNGGWGAYDVD

-continued

## Enzyme Sequences

NTRQLTNRIPPCNFGEVIDPPSEDEVTAHVLECFGSFGYDEAWKVIKRAVEYLKAQQRPDGSWF  
GRWGVNYYVYIGAVVPGGLKAVGVDMREPWWQKSLDWLVEHQNEGGWGE

>seq\_ID 278  
SPVWDTLLSLAMQDCERGFTPSMERALDWLANEVRYYGDSKKVGRVPEPSGWAFERANL  
NYPDIDDTAVALIVLARLPRAQLDQPRIREVIDRALGWTAMQSSNGGWAADFKNNDHLIITKIP  
FCDFGEALDPPSADVAHVLEALGLLGFERKHPAVERGLKFIKRSQEQEADGSWFGRWGVNHIY  
GTAAVLPALXAI GEDM

>seq\_ID 315  
SPVWDTALAVRALAAGLPPDHPFMTQATSWLLTQQIFKPGDWCIKCPDLPGGWAFXFHNN  
WYPDVDDSSMVLVALKDGDPDARHQAALQRGINWCLGMQSKNGGFASFDKDNKKEWLNAL  
PFGDLKALVDPPTEDITARILEMMGAFGHGLDHPDADRALAFRLRQHPPEGFWGRWGVNYYL  
YGTWSVLVALKRIGXDMSPYVQRAVNWIKSHQNPDGGWGEVCESTRHPPELGMQGPSTASQ  
TAWALLGLLAAGEIQAAEVKAGVDYLVKTKQNAQGRWDEKYFTGNWLP

>seq\_ID 297  
SPVWDTALVQLQAMQEASIPLDHPALVKAQWLLDREVRIKGDWIKSPGLEPGGWAFEFQND  
WYPDVDDSAAVLIAIKDIQVKNNKAKQGAARRGIDWCLGMQSKNGGWAADFKNKHLNKKIP  
FADLEALIDPPTADLTGRMLELMGNFGYDKHHPQAVHALEFLKKEQEPPEGFWGRWGVNHIY  
GTWYVLIGLEAIGEDMNQPYIKKAANWIKSRQNI DGGWGE

>seq\_ID 17  
QASISPVWDTGLAVLALRAAGLPADHDRLVKAGEWLLDRQITVPGDWVVKRPNLNPGGFALQF  
DNVYYPDVTAVVIALNLTLRLPDERRRRDAMTKGFRWIVGMQSSNGGWAADFVDNNTSDL  
PNHIFPCDFGEVTDPPSEDEVTAHVLECFGSFGYDDAWKVIQRAVAYLKRQKPDGSWFGRWG  
VNYIYGTGAVVSALKAVGIDMREPIQKALDWVEQHQNPDG

>seq\_ID 303  
SPVWDTALVVLVAMQEAGVPLDHPALVKAQWLLDREVRKGDWRKSPDIEPGGWAFEFLND  
WYPDVDDSGFVMLAIKDKVVRDKKKEQAIRKGINWCLGMQSSANGGWAADFKNTKYLLNKKIP  
PFADLEALIDPPTADLTGRMLELGLTFNFKDHHAIERALEFIIQLEQEPPEGFWGRWGVNHIY  
TWSVISGLEAIGEDMSQPYIRKTVNWLKSKQNMDGGWGE

>seq\_ID 298  
SPVWDTLVLVAMQEAGVPLDHPALVKAQWLLDREVRKGDWQVKS PDVEPGGWAFEFMN  
DWYYPDVTAVVIALXNIRVRDKKHEQAIRKGIWCLGMQSSANGGWAADFKNTKYLLNKKIP  
PFADLEALIDPPTADLTGRMLELMGNFDYSASYPAAVRALEFLKKEQEPPEGFWGRWGVNHIY  
GTWSVLCGLEAIGEDMSQPYIRKAVNWLKSKQNLDGGWGE

>seq\_ID 301  
SPVWDTALALVAMQEAGVPKDHPALVKAQWLLDREVRKGDWQIKSPELEPGGWAFEFLND  
WYPDVDDSGFVIMARDIKAPDKKHEQAIRKGIWCLGMQSSANGGWAADFKNTKYLLNKKIP  
FADLEALIDPPTADLTGRMLELMGSDYPMDHPPAAARALEFLKKEQEPPEGFWGRWGVNHIY  
GTWSVLCGLESIGEDMSQPYIKAVNWLKSKQNMDGGWGE

>seq\_ID 276  
SPVWDTLLTLLAMEDCDRGLTPSMQRALEWVLAQEVRYAGDWSKKVKGVEPSGWAFERANL  
NYPDIDDTAVALIVLARLPRAWLDEPRIRATIDRVLGWTAMQSSNGGWAADFKNDRPIITKIP  
FCDFGEALDPPSADVAHVLEALGLPGFDRRHPAVERGYKFLRSQEQEADGSWFGRWGVNHIY  
GTAAVLPALASIXEDM

>seq\_ID 283  
SPVWDTCLTSNALVESGGDTSAPHVHRSVQWLLNQEIRNHGDWSVKAPKVGPSGWAFEFAN  
KVYPDVDDAAEVIIALANYSNDSGTAPPDAIARGVRIWISGMQSSNGGWSFDKNNTSFFVTRL  
PFFDFGEVIDPPSVDVTAHVI EALAVAGWQEKASKQIQKALDYI WSEQEADGFWPGRWGINHIY  
GTCVLSALEAIGYDMADARVVKALKWIIECQNADGGWGE

>seq\_ID 307  
SPVWDTPWMI EALLETVGPPGDPALLRAGRWLMSKQITGVRGDWAMKSPKKGPGGWAFEF  
NDYYPDVTIQLVLTALCKLSIPWREKAKAVMQIDWLSIMQNDGGGWAADFDRNQTRWIVNRI  
PFDHKACLDPSDPDITGRMVEFLMRRNYS TSHPSVKKALKYI RETQEDFGAWFARWGINHIY  
GTWCVL TALAAMGIHTDSRVAKAVAWLSSVQRPDGGFSEADTYHPHKPFESYSESVPSSQ  
AWALMGLVAGGAVHSPAAARAACYLINNRNLNNGWDERHYTGTGFP

>seq\_ID 267  
SPVWDTAISVIALAESGLHRGHSVLVQATEWLVANEIRRGDQWQKNPTAPISGWAFEFKNDP  
YPDVDDTAMVLLALRHVHLYNDVVSQDREKSYLRGLNWLMSQCKNGGWAADFDRDNVKTIF  
EKIPFADHNAMIDPPSVDITGRVLELLGYVGYDKSYPCVTKALEYIKKQEQEADGSWYGRWGVN  
YIYGTWQVLRGLAAIGEDMQSEYVQKAVRWKSVQNP DGGWGE

>seq\_ID 309  
SPVWDTVLSITALADADLPRTHPAMRRAVAVLGLKQVLCBGDWRVKNRRGEPGGWSFEFNN  
NFYQNDDDTAAVLIAHKLARLPDEAKGEAMQRGLRWLLSMQCKDGGWAFDVRNKNRLLNKKIP  
PFADLESMLDPS TCDLTGRTEALGSI GFFP THRIVQHAVRFIRQHQEQEADGAWYGRWGVNHIY  
GTCHVLCGLLSVGEDMHQPYVQRAVQWLIHQNDADGGWGE

-continued

## Enzyme Sequences

>seq\_ID 202  
MVYSYEMMVLDDYPEDHPLRVECKAALKLVHRDDGSSYQCPCLSPVWDTAWSVMALEQA  
PSDARTETAIARAYDWLTDQVLDLRGDWENNAAPSTPPGGWAPQYENFYYPIDDSAVVLA  
MLHARGKRTGQPGRYEMPVARCLDWIIGLQSRNGGFGAFDANCRRDFLNAI PFADHGALLDP  
PTEDVSGRVLALGI TERPQDATARERCIQYLRDTQQPDGSSWWGRWGTNYI YGTWSVLAGLG  
LAGVDRKLPVNRGLQWLRGKQADGGWGETNDSYARPELAGKHEDGSMAEQTAWAMLG  
QMAVGEGDADSVHRGAAYLDAQNEDGFWMHYPHNAPGFPRIFHLKYHG

>seq\_ID 306  
SPVWDTPTVMALLEAGVPSNDPALLRSGRWLLAKQITDTKGDWAIKKNKNTAPGGWSFEPEN  
KYFPDVEDDTIEVLHCLHLKLAIPWREKEKPCRLGIDWLLSMQNDGGWGAFFDKNQKQVNVNRI P  
FSDHGACLDPPSSPDITGRMIEFLATQKFNSEYESVKRALKYIWKTEQDFGGWHARWGINYIYGT  
WCVL TGLRAIGFNM TDRRVQKALNWLESIQNKDGGFGESPASYEBCRYI PWKESVPSQTAWA  
LMALVAGGGAGSAPAENAATFLINRYNSNGVWDEECYTGTFP

>seq\_ID 281  
SPVWDTLLTLAYQDCELEMNDSAGRALDWILSQENSYRGDWAHRNKKLEPSGWAFERANLH  
YPIDDDTSVALIVLARLPQAVRSRPDIKSAIDRALAWTLGMQCRNGGWAFFDRDNDKLIITMIPF  
CDFSEALDPPSADVTAHVVEAMAHLGFDERSHKAVEKAYQYLLAEQEDDGSWFGRWGNHIY  
GTAAVLPALALGEDATVPHVKRAADWISAHQNTDGGWGE

>seq\_ID 310  
SPVWDTALAVRALAAGLPPEHPAMVKASEWLLTQQIFKPGDWSVKCPDLP PGGWAFEFVNN  
WYDPVDDSSMVLVALKEGLADAQKHAALQRGINWCLGMQSKNGGFASFDKDNTKEWLNAIP  
FGDLKALVDPPTEDI TARI LEMMGAFGHGLDHPVAVRGLAYLHQTRQRP EGPWWGRWGVNYIY  
GTWSVLVALKRI GEDMSRPHYVRRRAVDWVKAHQNP DGGWGE

>seq\_ID 311  
SPVWDTALAVRALAAGLPPEHPAMVKASEWLLTQQIFKPGDWSVKCPDLP PGGWAFEFVNN  
WYDPVDDSSMVLVALKGLVDAQKHAALQRGINWCLGMQSKNGGFASFDKDNTKEWLNAI  
PFGDLKALVDPPTEDI TARI LEMMGAFGHGLDHPVAVRALAYLHQTRQRP EGPWWGRWGVNYI  
YGTWSVLVALKRI GEDMNRPHYVRRRAVDWVKAHQNL DGGWGE

>seq\_ID 290  
SPIWDTAKAVNALHESGLPSDHPQLKAAARWLVEKEVRKPGDWKMRVPHVDVGGWPPQFRN  
EFYDPVDDTAAVVMALGRVDERDVPGIKDSITRGINWVTQMCCSCGGWAFFDVVKREFLT  
KVPYADHNAMLDPCCPDITGRCLEMYGRFPVVRKADVDVQRVIEKGI EYLKKTQEPDGSWYGRW  
GVNYIYGTWQSLKGLAAVGEDPSQPYIQKAAHFLKTHQNSDGGWGE

>seq\_ID 292  
SPVWDTAKAVNALHESGLPSDHPQLKAAARWLVEKEVRKPGDWKMRVPHVDVGGWPPQFR  
NEFYDPVDDTAAVVMALGRVDERDVPGIKDSITRGINWVTQMCCSCGGWAFFDVVKREFLT  
KVPYADHNAMLDPCCPDITGRCLEMYGRFPVVRKADVDVQRVIEKGI EYLKKTQEPDGSWYGR  
WGVNYIYGTWQSLKGLAAVGEDPSQPYIQKAAHFLKTHQNSDGGWGE

>seq\_ID 293  
SPVWDTCLSLAALTEAQAQNDHPAVKQAVEWLLDHQIFVEGDWCAQASGLEPGGWAFQYEN  
DKYDPVDDTGMVLMSLRAGVHDKHKKRVNVALNWLGMQNDPDSWGAFDI ENNYEYLN  
KIPFADHGALVDPGTADLTARCVELLAMLGYDATFPVVKRALEFLEHDQEDGWSYGRWGVN  
YIYGTWSVLCALGAIGEDVAKPYVRKSVQWLQDTQNE DGGWGE

>seq\_ID 313  
SPIWDTALAVRALTAAGMPPEHPAMVKASEWLLTQQIFKPGDWSVKCPDLP PGGWAFEFVNN  
WYDPVDDSSMVLVALKEGLADTAKHQAALQRGINWCLGMQSKNGGFASFDKDNTKEWLNAIP  
FGDLKALVDPPTEDI TARI LEMMGAFGHGLDHPVAVRALAYLHETQR PGGPWWGRWGVNYLY  
GTWSVLVALKRI GEDMSRPHYVRRRAVDWVKDQHNLDGGWGE

>seq\_ID 304  
SPVWDTPTVMALLEAGVPTDXPGLLRAGRWLISKQITGVHGDWAVKNRHALPGGWSFEFE  
NDYFPDVEDDTIEVLHVIHRLAIPWEEKSECCRLGLDWLLSMQNDGGWGAFFDRNQTLMVMNRI  
PFSHAACLDPPSPDIVGRVLEFLASRSFSREHPAVKRALDYIWRQSPFGGWWARWIDYLY  
GTWCVL TGLRAIGWDMEDPRVRKAWAWLESVARPDGGYGESPE SYRHSYVWKR SVPSQT  
AWALMGLVAGGVGHGAARGAADYLLTSRNAQGGWDEMDYTGTFP

>seq\_ID 291  
SPMWDTAKAVNALHESGLPSDHPQLKAAARWLVEKEVQKPGDWKMRVPHYVDVGGWPPQFR  
NEFYDPVDDTAAVVMALGRVDERDVPGIKDSITRGINWVTQMCCSCGGWAFFDVVKREFLT  
KVPYADHNAMLDPCCPDITGRCLEMYGRFPVVRKADVDVQRVIEKGI EYLKKTQEPDGSWYGR  
WGVNYIYGTWQSLKGLAAVGEDPSQPYIQKAAHFLKTHQNSDGGWGE

>seq\_ID 318  
SPVWDTGLALHALL ESGMDPDDPAIAKAMHWLDEREITDVAGDWAEQRPGLAPGGWAFQYR  
NDHYDPVDDTAVVGMAMHRANPQARPETLERTRAWIEGMQSQNGGWAFFDADNTHYHLNHI  
PFADHGAMLDPPTADVSARCLGMLSQMGYDRDHP SIQRAIAYLKNQDEEDGWSFGRWGTNYI  
YGTWSVLSALNAGEDMSQPYIRKAVDYLTNFRQREDGGWGE

-continued

## Enzyme Sequences

>seq\_ID 294  
 SPVWDTCLSLAALTEAGAQNDRHAPVAVKQAVWELLDHQIFVEGDWCDQAPGLEPFGGWAFQYEN  
 NKYPDVDDTGMVLMSSLRAGVHDKEHKRKRNVQALNWV LGMQNPDGSGWGFADIENNYEYLN  
 RIPPADHGALVDPGTADLTARCVELLAMLGYDATFPPVKRALEFLQDQEBEDGSWYGRWGVN  
 YIYGTWSVLCALGATGEDVAKPYVRKSVQWLQDTQNDGGWGE

>seq\_ID 320  
 SPVWDTCLGLHALL EAGEPREAPSVKKAVDWLLEREITETYGDVWVRRPHLKPSGWAFQYEW  
 NNYYPDVDDTAVVVMALDRVGDPRCRPAIERACEWIGMQSTSGGWGSDPENEFTYLNHIPP  
 ADHGALLDPPPTVDVTARCSMLAQVGYRHDHPAIRKSVXFILRQEEDKGSWYGRWGTNYVYG  
 TWSALSALNAVGEDMSSPVVRKGVAVLEAFQPDGGWGE

>seq\_ID 295  
 SPVWDTCLSLTAMTESGAHPEHPAVKQAVWELLDQQIFVKGDWADQAKNLEPFGGWAFQFEN  
 DRCPDVDDTGMVLMALLRAGVQDKEHKIKRINQAVNWV LGMQNPDGSGWGFADIGNDHEYLN  
 NIPPADHGALVDPGTADLTARCVELLAMLGYGDPFPIQRAVAFLEERDQEEFGAWYGRWGVN  
 YIYGTWSVLSAIGILGEDYAKPYVRKAVEWLKEIQNDGGWGE

>seq\_ID 324  
 SPVWDTSLAAHALL EAGEPNDPEVIGLLDWLKDQILTVGDWSARRPNLRPGGWAFQYENP  
 HYPDVDDTAVVAMAMHRQDQPKYAEAIARACEWLAGMQSSSGGWGAFDPENEHFYLNIPF  
 ADHGALLDPPPTVDVTARCVGCLAQVDAERFASEIQAGIDYIKRQEEDKGSWYGRWGANVYVYG  
 TWSALVALNKAGEDMNTPIYIRRAVDWLKARQRPDGGWGE

>seq\_ID 296  
 SPVWDTCLSLNALT EADMPANDPRVRAAVQWLFDRQIFVRGDWSENAPLEPFGGWAFQYEN  
 DKYPDVDDTGMVLMSSLRANAHEHDAQKRKNQALNWV LGMQNSDGSGWGFADIDNHYYTL  
 NNIPPADHGALVDPGTADLTGRCIELFGMLGYDKNFTPARRGIEFLKRDQHPCCGWYGRWGVN  
 NYLYGTWSVLTALGAIGEARDAPYLRRAVEWLYSVQNDGGWGE

>seq\_ID 305  
 SPVWDTPWMVALL EAGCPANDPXLIRAGRWLKAKXITEVRGDWAVKNRKPALPGGWSFEFE  
 NDYFPDVDDTIEVLSVIHRLSIPWNEKAKSCRLGLEWXLMSXNRDGGWGFADREQXFKVVNRI  
 PFSHAAALDPSDPDITGRMV EFLASXNFSKGHVAVRRALDYIWKQOAXFGGWAWARWIDYL  
 YGTWCVLTLGLASLGFXMDDPRARKAADWLESIQHADGGFGESESPYREDSFVDMKRSVPSQ  
 TAWALMGLVAAGRASGAAQRAAWLLDNRNNTNGSWDEQDYTGTFPP

>seq\_ID 282  
 SPMWDTSLAAHALMEADGRGDPKDNPRLISAMDWLADKQILDHVGDWAVRRPDVPPGGWAF  
 QYENPDYPDVDDTAVVVMAMHRADPERYEMSIDRACEWLVGMQSKNGGWGAFEPENEHY  
 LNSIPPADHGALLDPPPTVDVTARCVGALAQVDRDRYAAEIANGIRSI RREQEDDGSWYGRWGA  
 NYVYGTWSALVALKAGEDMQOPYIRRAVDWLKARQSDGGWGE

>seq\_ID 316  
 SPVWDTAWAVIGLCEGSMERTHPAVRSAIRWLYSMQILRPGDWAVKNPLTEPGGWAFEFHND  
 FYPDNDTAAVLMGLLFDLNDENHRAFERGVRWLLSMQNNDSGWGAFERNVDNKIFDQIP  
 FNDQKNMLDPSSTADVTGRVVELLGRIGRRLGGSFSEDPYVRQAI EFLKNEQEPEGCFGRW  
 VNYIYGTWSVLEAEIAGESMRAPYIRKAVNWVKKVQNPDDGGWGE

>seq\_ID 266  
 SPIWDTGIVLHSLVSESGVSPDHEALLRSVSWLLAKEVTHEGDWVKCPDAPVGGWYFEYANE  
 FNPDCDDTAKVLMATSRFSVDFPDAGRRLDARNRGLQWLLHMKNKDGWAAAFDQKCDNEL  
 LTYIPFADHNAMIDPSTEDITGRVLET LAREGFDNTHPVVKRAIQYLHKTQDAEGPWYGRWGS  
 NFIYGTWLVQLGLKAVGEDMTXPRYQRAANWLLNVQXNGSWGE

>seq\_ID 323  
 SPMWDTSLAAHAFLESGDREDPRLIRALDWLVKQILDHVGDWAVRRPGLRPGGWAFQYEN  
 PDYPDVDDTAVVAMAMHRTDPERYAENIDRACEWLAGMQSKNGGWGAFDPENEHYLNSIP  
 FADHGALLDPPPTVDVTARCI GCLAQVDAEAFADNIKRIGFIRKQEEDKGSWYGRWGANVYIYGT  
 WSALVALKAGEDMSQPYIRKSVAVLKGKRGQPDGGWGE

>seq\_ID 274  
 SPVWDTILSMQALLDTEKVFQPSPTLKKAMEWLLLEQQVRAWGDWVYVSDARGGGWAFQRA  
 NSFYPDVDDTINVMALRNVS PRGESKVVDEAIERALFWV LGMQCEGGWAAAFDRDNKAF L  
 TKVPPADHNAMIDPSTADLT SRTPEMFAMI APEVFTIHHPVVRGLEFLKDKQCKDGSWYGRW  
 GVNMYGTWQVLRGLRLIGEDMSKGYVRKGVWFKSVQLEDGGWGE

>seq\_ID 284  
 SPVWDTVAQLHALIASGLARRDEALRRAASWLLTRQSRTHGDWSGRNPAEPGGFYFEFRNEF  
 YPDVDDTAMALMVL TQAEANVATDVQHAARALAWMLGMQNRDGGWAAAFDRDNKHF L TQ  
 VPFADHNAMIDPSTADITGRV L GALSHPVS YGPDHPSVRRRAIAFLQRDQEPDGSWYGRWGVN  
 YLYGTQVLRGLRAIGFDMQQPFVRRRAARFLSAHQNDGGWGE

>seq\_ID 285  
 SPVWDTAITIALAESGLPKNHPAFEQAATWLEKKEIRFKGDWAVRMPGVEPESGWAFEHENKY  
 YPDTDDTMMVLMALRHVQSRNSAERCEQFDRALKWLLAFQCGDGGWAAAFDKDVTASWLEH

-continued

## Enzyme Sequences

VPFADHNAILDPTCSDLTARVLELLGSI SFDRQSAIVRRVAMMRRQTQETDGSWYGRWGVNYI  
YGTWQALRGLAAIGENMDQEWIRRRGRDWLESCQNDGCGWGE

>seq\_ID 308  
SPVWDTAIAAGYALGESGCAPQSSALRRMADWLLTKEVRRKDDWSVKRPDVEPSGWYFEFANE  
FYPD TDDTAMVLLSLLHGRATNPAAQEACAKRAVNWLLAMQSKDGGWAAFDVNDWKP LSY  
VFPADHNAMLDPSCPDITGRVLEALCKYGVSQEHPAVLRAIDYLIQTQE QDGSWHGRWGVNY  
VYGTFLALRGLKAAAGVSDREAYVLRAGEWLDLIQNP DGGWGE

>seq\_ID 288  
SPVWDTAIVAVSLAESGLEPDHPALQKSAEWLLDKEVRIQGDWAIKNRHGEASGWAFEFNNEF  
YPDVDDTLKVVLLALRLIKTRDEETKREAMERALGWVMSFQCSDGGWAAFDKDV TQRWLEDVP  
FADHNAILDPTCSDI TARCLELLGKMGCTSDHPAVRRALRMVRETQE PDGTWWRWGVNYIY  
GTWQILRGLSALKIDMNQDWI VRAKEWLESCQNP DGGWGE

>seq\_ID 287  
SPVWDTAITSVALTSSGVKPDHPQIQKAADWLLDREVVMRGDWKVKNPYPHAGSWAFEFNND  
FYPDADDTFKVVLLALMKMKSSDPERQKIMDRALDWAR SFQCKDGGFAAFDKDVTKKWLEHV  
PFADHNAILDPS CSDITARGLECMGKLGWRPTDRVIRRAIRYLKKTQE EDGSSWGRWGVNYIY  
GTWQSLRGLLEAIGEDMNQDWVVRARNWLESCQNP DGGWGE

>seq\_ID 289  
SPIWDTAIVTMAIAESGQDPNDPRLQKAADWLLEREIGFRGDWRENCDFPEATGWAFEFNND  
WYPDVDDTFQVILGLKPLSASDSRRQEQLDRAIPWC RAMQCREGGFAAFDKINDAWLNEV  
PFADHNAILDPPCSDI TGRALETLSLMGFDRREDPVVRRARQYLMETQLEDGSSWFRWGVNYIY  
GTGHALRGLHAI GEDINGSAMQARARNWLENCQNDGCGWGE

>seq\_ID 286  
SPVWDTAINVISLAESGLLSDHPALQKAADWLVNKEVFRFGDWSVNNSPYQVSGWAFYNNV  
YYPD TDDTAMVLMALRLIRPKDPQALNELFRRALDWQLSFQCRDGGWAAFDKNTT PWLEDM  
PFADHNAILDPTCSDLTARTLELLGYTFDPKQASVRDALQYLIDTQEDGDSWYGRWGVNYIY  
GTWQVLRGLRAMGQDMTQDWILRGRDWLESCQNSDGGWGE

>seq\_ID 270  
SPVWDTALAMSALLEGD TAPDDEALQRGCRWLLGKEVRRHRGDWQVNVGAEPGGWFFEYEN  
EFYPCDDTAEVLAVLERVRLSDPEEQRRRDLDRALAWQLGMQSTNGGWGAFDKDCDHR  
ILELVPFADHNAMIDPPTVDVTSRSIEAALAMGVPASDAAIRRAVRFLYSEQEADGSWYGRWG  
SNLYGTWLALCALRSAGEDLTPAVQRAVEWLLSVQQEDGCGWGE

>seq\_ID 322  
SPVWDTGIAAHALGEAGHASAMQSTADWLLTKEVRRKGDWSVKRPDVEPSGWYFEFANEFY  
PDIDDTAQVLLGLAHAKASDPAKQKACMDRAVAWLLAMQSGDGGWAAFDVNDNWEFLSVP  
FADHNAMLDP TPCPDITGRVLEALAACGVPNSHPAVKRGVEFLRNSVEKDGSWYGRWGVNYIY  
GTYLALRGLRASGEDDREAHLRAGEWLRAIQNADGCGWGE

>seq\_ID 263  
SPVWDTSLILNALLAGSEKTE TDPKILKAGQWLLDREVREIGDWIKNRRGVPVGGWYFEYANE  
FYPCDDTAEVITV LNQMFS DPEKEKAKQVAQQRGLDWLLSMQNKDGGWPAFDKNC DKQS  
LTYMPFADHNAMIDPSYEDITGR TLEALASLGFSEDDPIVRRAVDFLKSQLPDGTWYGRWGC  
NFLYGTWLAI SGLYHAGEDLNERYQSLLSWLEQCQNE DGGWGE

>seq\_ID 268  
SPVWDTCLILNSMLEHLEPDHPRVQKAAEWLLSKEVTEPGDWQVKCPEAPVGGWYFEYANEF  
YPCDDTAEVLAALQRVQFTDADREAQKRGAIQRGLGWLLAMQNQDGGXAAFDRECTREALT  
YVPFADHNAMIDPSNGDITGRV LKALDYAGYSPDDPIVRGGVDFLLANQEPDGTWYGRWGCN  
HLYGSWLAVVWGLKHAGVNLQQTQFTQVM SWLESCQNDGCGWGE

>seq\_ID 265  
SPVWDTTNAMTAVLDAGLPGNHPAVLRAARWLLSKEVRMPGDWRLWYKNGE PGGWFFEYFN  
NEFYPDADDTAEALHCLCRVVFDCEDMDRCRAAIKRG LNWQFACQNP DGGWPAFDKECDD  
EYLTFFIPFADHNAMIDPSCCDITGRSLQALS KLG YTTNDVDVKRAIDYLLDAQEDDGTWYGRWG  
INYIYGTWLAVQGLRAIGVDLSEKRFQKVTKWL RKKQNP DGGWGE

>seq\_ID 269  
SPVWDTCLILNSLLEHLEPDHPRLQHAAEWLLSKEVTEPGDWQVKCPEAPIGGWYFEYANEFY  
PDCDDTAEVLAALQRVRFSDADREAQKHAAIERGLGWLLAMQNQDGGWAAFDRECTREALT  
YVPFADHNAMIDPSNGDITGRV LKALDYSGRSPQDPVVGQGVHFLLANQEPDGTWYGRWGC  
NHLYGSWLAIWGLKHAGVDSQQSQFMRLLSWLESCQNP DGGWGE

>seq\_ID 319  
SPVWDTLSAHALMEAGLEENDKRL EGLLDWLKDLQILDVKGDWVARRPDV RGGWAFQYR  
NDHYPDVDDTAVVAMAMHRQGEKYKEAIDRAAEWI VGMQSSGGWGA FDPENEHFYLNLSI  
PFADHGALLDPPTEDVTARCVGFLAQLD PDAYAEPIKRGVEFLKRTQEQEDGSSWGRWGANF  
VYGTWSVLCALNAAGEDPKSPYIQKAVAWLKS RQREDGCGWGE

-continued

Enzyme Sequences

```
>seq_ID 321
SPVWDTGIACQALQEVGGPAADAGVQRALDVLVERQLRDEPGDWRDRPDLEGGGWAFQY
NNPHYPDLDDTSMVAWVMQVADHGRYREEIRRAAKWVVGMRSEGGGFASFEVDNTYYLNLH
IPFADHGKLLDPPTXDV TARCIAVLAI TDRAQHE TVIREAIDFLFVDQEBEGSWFGRWGTDYIYG
TWSVLSXLDVVGDFMRDARVRXSVEWLFXQQNPDGGWGE

>seq_ID 272
SPVWDTGLVALALQEVQDKHNSQDALQRNLKQAYSWSLKSQKLDKDEPGDWRISKPTLTGGGWAF
QFNNPHYPDVDDTAVVAFALAQAEHTELDESIHLATRWIEGMQSQNGGYGAFDNDNTFYLLNE
IPFADHGALLDPPTADVSARCAML MARVAKDHEEYLPALERTIQYLRSEQEBADGSGWFGRWGT
NYVYGTWSVLLGLEQTNVPKTDPLFTKAAQWLKSVQRPDGGWGE

>seq_ID 273
SPVWDTGLVALALPEVDKHN SQDALQPNLKQAYSWSLKSQKLDQPGDWRISKPTLTGGGWAF
QFNNPHYPDVHDTAVLAFALAQAEHTELDESIHLATRWIEGMQSQNGGYGAFDNDNTFYLLNE
IPFADHGALLDPPTADVSARCAML MARVAKGHEEYLPALERTIQYLRSEQEBADGSGWFGRWGT
NYVYGTWSVLLGLEQTNVPKTDPLFTKAAQWLKSVQRPDGGWGE

>seq_ID 317
SPVWDTILGMI GLVDCGHGDKDPLLV TARDWIVKRQLLVNYGDWKVYVNPNGPSGGWSFEYDN
SWYPDVDDTAAIIVIGFLKQDYEF RHSEVVKRACDWIASMQNXGGWAAPDINNDKTFLENIIP
SDMESLCDPSSPDVVGRVLEAFGILNDPKYAEVCRREGI EYLRRTQES EGSWFGRWGVNYVYG
TSNVLC SLKRQDVAXKDP MVTRAL TWLKKVQNKDGGWGE

>seq_ID 215
MGRQTRNL TRREPAEAEERGFRL LD AHRRADSSWV GELSSSALATAMSALALRLLGHPAES
GPVAGGLAWLAATRNPDGGWDAPGEP SNMNATSIAAALARCAPRRYREEVAGRRRWVE
EHGGFAALNDPRTTTL SGPGR TLWALAGLVPPERVKLPTEMI LLPRRIRRTVSTTFPAFLSLSL
LHERFRPSRWRRLRRRAER EALAWLRRRQQGPNGSYEESAF L TSLIAAAL TAAGAEGD IVR
RALPFVLR SRRPDGSWP IDRDLENFDTTQAI LAHHEAGRPLREAGRVREWLLDNQFRFPFPPT
SPPGGWAWAYPAGWPD TDDTACALRS LRL LGVPAGHPS IRLGLRWLYRMQNRDGSWPTFV
RGSRMPFDHGCPYI TSQVLSALALMGPEARRGAPLRRALAYLRRRQRPDGS LGSLWFRPHTR
GTAAAVEAFSDLGLSGDPLVGRAARWLAEHQNP DGGWGDGHGAPSTABETAWAS AALLRLG
GGEAARKGVRVLEHQDPGGWKPAVIGLYYASLSYS DTFYALSYP LVALARHRLSR

>seq_ID 191
MIKKILV LILLMVVTSKVDI ERVQTV IRDAREI CWNELTDNEWVYPTYLGLTFLSEYFELKALGI
QNSQFEESKFTQ ILLGSQLPDGSWVQVEDAYIQTGQLD ATIFNYWYLKAVGDIHTDTMKKAQE
WIKANGGI EKAQTMTKFKLAMPGQY PWKKL FKIP LI LFYKKNPL YIKDIT AQWVYPHMTALAYL
QNQR I I FNVAVS IS ELYKNKAPKI KNHQKGRPS FF INNLVQEM LKLRQPMGS FGGYTVS TLLSM
LALNDY TGR TNKHKSEI SDAL KGLDFVEFN YFNFRQAYHGS LDDGRWDT L I ISWAMLESGE
DKEKVRPI VENMLQKGVQPNNGGIEYGYDFGYAPDADD TGLLLQVLSY YGTDYADAMDKGAEF
VYSVQNTDGGFPAPDKGMGNPLYKYAFKIAGI ADSAEI FDPSSPDVTAHILEGLISSDRSNYD
VVVKS LKYFMDTQENFGSWEGRWGINIYIYAAGAVLPAL KKMNGWAKAVNWLVS KQNADGG
FGETT LSYRDPK KYNGI GVSTVTQTSWGLLGLLAVEDHYDVKEAI EKARDGEFKDISV VGTGHR
GLLYLQYPSYARSFPV I SLGRFLDQQR
```

SEQUENCE LISTING

The patent contains a lengthy “Sequence Listing” section. A copy of the “Sequence Listing” is available in electronic form from the USPTO web site (<http://seqdata.uspto.gov/?pageRequest=docDetail&DocID=US09447404B2>). An electronic copy of the “Sequence Listing” will also be available from the USPTO upon request and payment of the fee set forth in 37 CFR 1.19(b)(3).

55

The invention claimed is:

1. An enzyme mutant with cyclase activity which is a mutant of a wild-type enzyme comprising the amino acid sequence of SEQ ID NO: 2 with a mutation at a position corresponding to position F486 of the amino acid sequence of SEQ ID NO: 2, wherein up to 10% of the amino acid residues in said enzyme mutant are altered relative to the amino acid sequence of SEQ ID NO: 2 by deletion, insertion, substitution, addition, inversion, or a combination thereof, and wherein said enzyme mutant catalyzes at least the cyclization of a citronellal isomer to at least one isopulegol isomer.

60

65

2. The enzyme mutant of claim 1, wherein up to 5% of the amino acid residues in said enzyme mutant are altered relative to the amino acid sequence of SEQ ID NO: 2 by deletion, insertion, substitution, addition, inversion, or a combination thereof.

3. The enzyme mutant of claim 1, wherein the mutation at the position corresponding to position F486 of the amino acid sequence of SEQ ID NO: 2 is a substitution selected from the group consisting of F486N, F486Q, F486L, F486M, F486E, F486G, F486S, F486V, F486T, F486C, F486I and F486A.

## 165

4. The enzyme mutant of claim 1, wherein said enzyme mutant further comprises at least one mutation at a position corresponding to position W374, D437, D440, F428, W555, Y561, Y702, or Y705 of the amino acid sequence of SEQ ID NO: 2.

5. The enzyme mutant of claim 1, wherein said enzyme mutant does not comprise a mutation at a position corresponding to position D437, D439 and/or D440 of the amino acid sequence of SEQ ID NO: 2.

6. The enzyme mutant of claim 1, wherein said enzyme mutant does not comprise a mutation at a position corresponding to position Y702 of the amino acid sequence of SEQ ID NO: 2.

7. The enzyme mutant of claim 1, wherein said enzyme mutant further comprises at least one mutation at a position corresponding to position P229, D439, D508, E601, G553, G556, N432, P436, P499, R224, S371, T376, T563, W414, or W624 of the amino acid sequence of SEQ ID NO: 2.

8. The enzyme mutant of claim 1, wherein said enzyme mutant is:

a) a single mutant comprising F486X, with X=N, Q, L, M, E, G, S, V, T, C, I, or A, of the amino acid sequence of SEQ ID NO: 2;

or

b) a multiple mutant comprising F486A/Y702A, F486A/Y561A, or F486A/Y705A of the amino acid sequence of SEQ ID NO: 2.

9. The enzyme mutant of claim 1, wherein said enzyme mutant displays at least 50% of the citronellal-isopulegol

## 166

cyclase activity of an enzyme that comprises the amino acid sequence of SEQ ID NO: 2 from position 1 to 725, from position 2 to 725, or from position 16 to 725, and optionally N-terminally extended with a methionine residue.

10. The enzyme mutant of claim 9, wherein the citronellal-isopulegol cyclase activity is determined using a citronellal as a reference substrate under standard conditions.

11. The enzyme mutant of claim 1, wherein the mutation takes place in an enzyme that comprises the amino acid sequence of SEQ ID NO: 2 from position 1 to 725, from position 2 to 725, or from position 16 to 725, optionally extended N-terminally with a methionine residue.

12. The enzyme mutant of claim 1, wherein said enzyme mutant comprises a single mutation at a position corresponding to position F486 of the amino acid sequence of SEQ ID NO: 2.

13. The enzyme mutant of claim 1, wherein said enzyme mutant comprises an additional mutation at a position corresponding to position Y702 of the amino acid sequence of SEQ ID NO: 2, and wherein said mutation is a Y702F substitution.

14. The enzyme mutant of claim 1, wherein said enzyme mutant further comprises a mutation selected from the group consisting of:

Y702X, with X=F, A, C, or S, of the amino acid sequence of SEQ ID NO: 2; and

Y561X, with X=A or S, of the amino acid sequence of SEQ ID NO: 2.

\* \* \* \* \*